

MASTER'S THESIS

De Effectiviteit van Verschillende Feedbackvormen op de Leerresultaten van een Serious Game in het Basisonderwijs

Kouwenhoven, Maike

Award date:
2019

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

Take down policy

If you believe that this document breaches copyright please contact us at:

pure-support@ou.nl

providing details and we will investigate your claim.

Downloaded from <https://research.ou.nl/> on date: 05. May. 2023

Open Universiteit
www.ou.nl



**De Effectiviteit van Verschillende Feedbackvormen op de Leerresultaten van een
Serious Game in het Basisonderwijs**

**The Effectiveness of Different Types of Feedback on Learning Outcomes of a Serious
Game in Primary Education**

Maike Kouwenhoven

Master Onderwijswetenschappen
Open Universiteit

Datum: 29 augustus 2019
Begeleider: Dr. Rob Nadolski

Inhoudsopgave

Samenvatting	3
Summary	5
Inleiding	7
Vraagstellingen en hypothesen	14
Pilot	15
Methode	18
Resultaten	23
Conclusie en discussie	27
Referentie	33
Bijlage 1. Verantwoording serious game	36
Bijlage 2. Berekening G*Power	43
Bijlage 3. Kwaliteitsmeting game	44
Bijlage 4. Kennistesten	49

Samenvatting

De Effectiviteit van Verschillende Feedbackvormen op de Leerresultaten van een Serious Game in het Basisonderwijs.

Maike Kouwenhoven

Serious games, educatieve games waarbij leren centraal staat, worden steeds vaker toegepast in het onderwijs. Om de leereffecten van deze serious games te bevorderen, is het belangrijk dat leerlingen feedback krijgen op hun handelingen binnen het spel. Onderzoekers maken onderscheid tussen drie vormen van feedback: *knowledge of results* (KR), *knowledge of correct response* (KCR) en *elaborated feedback* (EF). Uit onderzoek in het voortgezet- en hoger onderwijs blijkt dat KR feedback het minst effectief is voor het verbeteren van de leerresultaten van leerlingen, KCR feedback kan effectief zijn wanneer het om eenvoudige denkniveaus gaat en EF feedback is ook bij meer complexe denkniveaus effectief.

De vraag is of deze resultaten ook voor het basisonderwijs gelden. Doel van dit onderzoek is daarom te meten hoe effectief de feedbackvormen KR, KCR en EF zijn voor het verbeteren van de leerresultaten van leerlingen op de basisschool bij het spelen van een door de onderzoeker ontworpen en ontwikkelde serious game: Plantendetective. De verwachting is dat KCR en EF ook voor leerlingen in het basisonderwijs effectiever zijn dan KR en dat EF daarnaast ook effectiever is dan KCR. Om de kwaliteit van de serious game te kunnen waarborgen, is bij het ontwerpproces gebruik gemaakt van het framework van Annetta (2010). Het prototype van de game Plantendetective is in een pilot getest en geëvalueerd door een groep van drie leerlingen die willekeurig geselecteerd zijn uit alle participanten. Tijdens een *cognitive walkthrough* speelden deze leerlingen Plantendetective terwijl zij hardop vertelden wat zij dachten. Vervolgens beoordeelden zij de zes elementen van Annetta (2010) door middel van een kwaliteitsmeting. Op basis van de bevindingen uit deze pilot is de game aangepast voor gebruik in het interventieonderzoek.

Het interventieonderzoek heeft een gerandomiseerd pretest-posttest-retentietest ontwerp gevolgd met twee experimentele groepen (KCR en EF) en een controlegroep (KR). De deelnemende leerlingen zijn willekeurig over de drie onderzoeksgroepen verdeeld (N = 140). Alle leerlingen startten met een voormeting waarbij hun voorkennis op het gebied van natuuronderwijs is gemeten. Vervolgens speelden de leerlingen Plantendetective waarbij zij, afhankelijk van de onderzoeksgroep, KR, KCR of EF feedback kregen aangeboden. Direct na afloop van de game vulden de leerlingen een kwaliteitsmeting in om de kwaliteit van Plantendetective te beoordelen. Diezelfde middag is een nameting afgenomen. Na twee weken vond op dezelfde wijze een retentiemeting plaats.

Voor de voor-, na- en retentiemeting is gebruik gemaakt van drie kennistesten met tien meerkeuzevragen die door de onderzoeker zelf zijn ontwikkeld. De kwaliteitsmeting is samengesteld

uit de Intrinsic Motivation Inventory (Ryan & Deci, 2000), de Flow-short (Rheinberg, Vollmeyer, & Engeser, 2003), de NASA-TLX (Hart & Staveland, 1988), studeerbaarheidsvragen (Nadolski & Hummel, 2017) en nog een aantal aanvullende vragen.

Uit de resultaten van dit onderzoek blijkt dat de game Plantendetective kwalitatief voldoende is en bovendien een significant leereffect veroorzaakt bij de leerlingen. De gemiddelde scores op de nameting en retentiemeting zijn echter niet significant verschillend voor de drie onderzoeksgroepen. Op basis van dit onderzoek kan dan ook niet geconcludeerd worden dat één van de drie feedbackvormen effectiever is voor leerlingen in het basisonderwijs dan de anderen. Vervolgonderzoek zal moeten uitwijzen of deze bevinding ook geldt bij een grotere groep participanten en verbeterde meetinstrumenten.

Trefwoorden: Serious games, feedback, game-based learning, basisonderwijs

Summary

The Effectiveness of Different Types of Feedback on Learning Outcomes of a Serious Game in Primary Education

Maike Kouwenhoven

Serious games, educational games with learning as the main purpose, are increasingly being used in education. To improve the learning effects of these serious games, it is important that pupils receive feedback on their actions during the game. Researchers distinguish three types of feedback: knowledge of results (KR), knowledge of correct response (KCR) and elaborated feedback (EF). Research in secondary and higher education shows that KR feedback is the least effective for improving learning outcomes. KCR feedback can be effective when it comes to lower learning outcomes and EF feedback is also effective for more complex learning outcomes.

However, it is unclear whether these results also apply to primary education. Therefore the aim of this study is to measure how effective KR, KCR and EF feedback is for improving the learning outcomes of primary school pupils after playing a serious game designed and developed by the researcher: Plant Detective. It is expected that KCR and EF are more effective than KR for primary school pupils and that EF is more effective than KCR. In order to guarantee the quality of this serious game, the framework of Annetta (2010) is used in the design procedure. The prototype of the game Plant Detective has been tested and evaluated in a pilot that existed of three students who were randomly selected from all participants. These students performed a cognitive walkthrough by playing Plant Detective while thinking out loud. They assessed the quality of the game according to the six elements of Annetta (2010). Based on the findings from the pilot, the game Plant Detective has been adapted to be used as an intervention in an experimental research design.

This study followed a randomized pre-test-post-test-follow-up-test design with two experimental groups (KCR and EF) and a control group (KR). The participating pupils were divided over the three research groups (N = 140). All pupils conducted a pre-test in order to measure their prior knowledge on the subject of biology. The pupils played the game Plant Detective afterwards and, depending on their research group, received feedback in form of KR, KCR or EF. Pupils completed a quality test immediately after they finished playing their game, in order to assess the quality of Plant Detective. A post-test was taken that same afternoon. After two weeks a follow-up test took place in the same way.

For the pre-test, post-test and follow-up test, three knowledge tests with multiple choice questions were used, developed by the researcher. The quality test was a combination of the Intrinsic Motivation Inventory (Ryan and Deci, 2000), the Flow short (Rheinberg, Vollmeyer, & Engeser,

2003), the NASA-TLX (Hart & Staveland, 1988), feasibility questions (Nadolski & Hummel, 2017) and a number of additional questions.

The results of this study show that the quality of the game Plant Detective is sufficient and causes a significant learning effect. However, there is no difference in the average scores on the post-test and follow-up test of the three research groups. Based on this study it cannot be concluded that one of the three types of feedback is more effective for pupils in primary educations than the others. Further research will have to show whether this finding also applies with a larger group of participants and improved measurement tools.

Keywords: Serious games, feedback, game-based learning, primary education

Inleiding

De generatie kinderen die nu naar school gaat, is geboren in een tijdperk dat in het teken staat van technologie. Deze ‘Net Generation’ of ‘Digital Natives’ zijn gewend aan het gebruik van allerlei soorten software en entertainment games (Girard, Ecalle, & Magnan, 2013). In het onderwijs worden spelvormen al sinds lange tijd toegepast voor de ontwikkeling van kinderen. Door middel van spel kunnen kinderen namelijk cognitieve, affectieve en motorische vaardigheden aanleren (Lai, Ang, Por, & Liew, 2018). De groeiende populariteit van entertainment games heeft echter ook geleid tot een toenemende interesse in educatieve games (Plass, Homer, & Kinzer, 2015). Deze educatieve games, waarbij leren centraal staat, worden ook wel serious games genoemd (Papanastasiou, Drigas, & Skianis, 2017). Uit onderzoek blijkt dat serious games tot betere leerresultaten *kunnen* leiden dan de traditionele instructie (D. B. Clark, Tanner-Smith, & Killingsworth, 2016; Wouters, Nimwegen, Oostendorp, & Spek, 2013). Echter is dit bij lang niet alle serious games het geval (Annetta, Minogue, Holmes, & Cheng, 2009; R. E. Clark, Yates, Early, & Moulton, 2010; Girard et al., 2013; Wrzesien & Alcañiz Raya, 2010). Deze verbeterde leerresultaten kunnen bijvoorbeeld uitblijven doordat de instructie onvoldoende geïntegreerd is in de serious game (R. E. Clark et al., 2010). Om de leereffecten van serious games te bevorderen, is het cruciaal dat leerlingen effectieve feedback krijgen op hun handelingen binnen het spel (Bellotti, Kapralos, Lee, Moreno-Ger, & Berta, 2013). Effectieve feedback wordt precies op het juiste moment geboden, is frequent genoeg en bevat de juiste informatie om een leerling verder te helpen. Deze feedback kan een brug slaan tussen de kennis waarover een leerling beschikt en de kennis waarover hij zou moeten beschikken om het gestelde leerdoel te behalen (Hattie & Timperley, 2007). Niet alle vormen van feedback zorgen echter voor een even effectieve en efficiënte verwerving van het leerdoel. Onderzoekers maken grofweg onderscheid tussen *knowledge of results* (KR), *knowledge of correct response* (KCR) en *elaborated feedback* (EF) (Shute, 2008). KR geeft alleen aan of het antwoord goed of fout is, KCR laat daarnaast ook zien wat het juiste antwoord is en bij EF krijgt de leerling aanvullende informatie, bijvoorbeeld over het gegeven antwoord of het onderwerp van de les (Tsai, Tsai, & Lin, 2015). Uit onderzoek blijkt dat EF de meest effectieve vorm van feedback is om de leereffecten bij serious gaming te vergroten (Jaehnig & Miller, 2007; Mason & Bruning, 2001; van der Kleij, Feskens, & Eggen, 2015). De vraag is echter of EF ook het meest effectief is voor leerlingen in het basisonderwijs. Naar deze doelgroep is namelijk nauwelijks onderzoek gedaan (van der Kleij et al., 2015). Dat onderzoek naar serious games zich voornamelijk op het voortgezet- en hoger onderwijs richt, is een bekend probleem (Abdul Jabbar & Felicia, 2015; Li & Tsai, 2013; Perrotta, Featherstone, Aston, & Houghton, 2013). Volgens van der Kleij et al. (2015) laten de zeldzame onderzoeken die beschikbaar zijn kleinere positieve leereffecten voor EF in het basisonderwijs zien dan in het hoger onderwijs. Dit suggereert volgens hen dat feedbackmechanismen

wellicht anders werken bij kinderen op de basisschool. Daarom is de vraag wat binnen het basisonderwijs de effectiviteit van verschillende feedbackvormen is voor de kennistoename van leerlingen bij serious gaming. Hierbij zijn vooral de effecten van KCR en EF interessant, omdat uit onderzoek blijkt dat KR het minste effect heeft op de leerresultaten (van der Kleij, Timmers, & Eggen, 2011). *Response contingent* EF, een variant van EF die uitlegt waarom het incorrecte antwoord fout is of waarom het correcte antwoord goed is, zou de meest effectieve vorm van EF zijn voor leerlingen in het basisonderwijs die over weinig voorkennis beschikken (Mason & Bruning, 2001). Om de verschillen in effectiviteit van KCR en *response contingent* EF te onderzoeken, is binnen dit onderzoek een eigen serious game ontworpen door de onderzoeker: Plantendetective. Om de kwaliteit van deze serious game te kunnen waarborgen, is bij het ontwerp gebruik gemaakt van het framework van Annetta (2010) en is de hulp ingeschakeld van *subject matter experts*. Het prototype van de game Plantendetective is bovendien geëvalueerd door een pilotgroep en op basis van deze evaluatie aangepast voor gebruik binnen het interventieonderzoek.

Definitie feedback

Voor het verbeteren van de leerresultaten van leerlingen is het van groot belang dat feedback wordt gegeven en dat deze feedback op de juiste wijze wordt toegepast (Shute, 2008). Om inzicht te krijgen in de manier waarop feedback kan worden ingezet, moet het begrip eerst beter worden gedefinieerd. Feedback is een reactie op een actie van een leerling met als doel het denken of het gedrag van de leerling aan te passen en zo het leren te bevorderen (Mason & Bruning, 2001; Shute, 2008). Deze reactie bestaat uit informatie die beschikbaar wordt gesteld door een medium, dit kan bijvoorbeeld een leerkracht of een medeleerling zijn, maar ook een boek of een digitaal programma (Hattie & Timperley, 2007). Het belangrijkste doel van feedback is het verkleinen van het verschil tussen de kennis waarover de leerling op dat moment beschikt en het leerdoel dat hij uiteindelijk moet behalen (Hattie & Timperley, 2007). Dit verschil kan worden verkleind doordat feedback de leerling helpt met het identificeren en verbeteren van fouten en misconcepties (van der Kleij et al., 2015). Door middel van feedback krijgen leerlingen de informatie en hulpmiddelen die ze nodig hebben om te slagen (Rogers, 2017). Hierdoor is een leerling in staat zijn kennis en vaardigheden te vergroten (Shute, 2008).

Soorten feedback

Volgens Hattie en Timperley (2007) kan feedback worden ingedeeld in vier niveaus. Het eerste niveau betreft feedback die gericht is op de taak zelf, dit geeft aan of de opdracht correct of incorrect is uitgevoerd. Het tweede niveau gaat in op het proces rondom de uitvoering van de opdracht, waarbij bijvoorbeeld gefocust wordt op het verwerken van informatie en het gebruik van strategieën. Bij het derde niveau staat zelfregulatie centraal. Deze feedback heeft invloed op de manier waarop leerlingen

naar zichzelf kijken en hun eigen handelingen evalueren. Tot slot richt het vierde niveau van feedback zich op de leerling zelf. Deze feedback is niet gerelateerd aan de prestatie, maar kan bijvoorbeeld opmerkingen betreffen zoals “Je bent een goede leerling”.

Naast de verschillende niveaus van feedback, wordt binnen de literatuur ook onderscheid gemaakt tussen drie verschillende categorieën: *Knowledge of results* (KR), *Knowledge of Correct Response* (KCR) en *Elaborated Feedback* (EF). KR of verificatiefeedback is de meest simpele vorm waarbij de leerling alleen te horen krijgt of het antwoord goed of fout is. Bij KCR krijgt de leerling hiernaast ook het goede antwoord aangewezen. Deze feedback is evenals KR alleen gericht op de taak. EF kan zich richten op alle vier de levels van feedback (van der Kleij, Eggen, Timmers, & Veldkamp, 2012). Dit is dan ook de meest uitgebreide vorm van feedback die bestaat uit KR of KCR aangevuld met extra informatie (Tsai et al., 2015; van der Kleij et al., 2015). De strekking van deze informatie is afhankelijk van de variant van EF die wordt toegepast. Shute (2008) maakt onderscheid tussen zes varianten die in oplopende mate informatie bieden aan de leerling. De eerste vorm is *attribute isolation* waarbij informatie wordt gegeven over de belangrijkste kenmerken van het concept dat de leerling moet oefenen. Bij *topic contingent* feedback krijgt de leerling informatie over het onderwerp van de les, bijvoorbeeld door middel van extra instructie. *Response contingent* feedback richt zich op het antwoord van de leerling en beschrijft waarom het incorrecte antwoord fout is of waarom het correcte antwoord goed is. Een volgende variant van EF is het geven van hints om de leerling in de juiste richting te wijzen, bijvoorbeeld in de vorm van een uitgewerkt voorbeeld of een demonstratie. Bij *bugs / misconceptions* krijgt de leerling informatie over specifieke fouten en misconcepties. De meest uitgebreide vorm van EF is *informative tutoring*. Deze feedbackvorm omvat verificatie, het uitlichten van fouten en het geven van hints over het vervolgproces. Naast KR, KCR en EF classificeert Shute (2008) *try again* als feedbackvorm. Hierbij mag de leerling één of meerdere keren proberen zijn antwoord te verbeteren. Van der Kleij et al. (2015) zien *try again* echter niet als aparte vorm van feedback, maar als aanvulling die gebruikt kan worden in combinatie met KR of EF.

Feedback kan tot slot worden ingedeeld op grond van timing. Directe feedback biedt meteen na het afronden van de taak informatie aan de leerling, terwijl uitgestelde feedback pas later beschikbaar komt. Dit kan enkele minuten, uren of zelfs weken later zijn (Shute, 2008).

Effectiviteit van feedback

Niet alle vormen van feedback zijn even effectief als het gaat om het verbeteren van de leerresultaten (Hattie & Timperley, 2007; Shute, 2008). Het is echter niet mogelijk om een feedbackvorm aan te wijzen die universeel het beste werkt, daarom is het belangrijk om voor alle schoolniveaus en leerdoelen te onderzoeken wat de beste resultaten oplevert (Mason & Bruning, 2001). Naar aanleiding van eerder onderzoek kan wel een aantal algemene conclusies worden getrokken. Zo geven Hattie en

Timperley (2007) aan dat feedback gericht op het proces of zelfregulatie krachtig is wanneer het gaat om het diepere proces van leren en het beheersen van een bepaalde vaardigheid. Taakgerichte feedback is vooral effectief wanneer sprake is van een verkeerde interpretatie van de leerstof. Feedback gericht op de leerling zelf is het minst effectief. De serious game die binnen dit onderzoek centraal staat, is gericht op het aanleren van concepten en procedures binnen de lesstof. Daarom ligt de focus op taakgerichte feedback.

Wat betreft de categorieën van feedback kan gesteld worden dat KR meestal niet geschikt is om de leerresultaten van leerlingen te verbeteren (Tsai et al., 2015). In hun meta-analyse concluderen van der Kleij et al. (2015) namelijk dat KR een zeer klein effect heeft op de leerresultaten. Wanneer het leerdoel van de les zich op eenvoudige denkniveaus richt, zoals het herkennen of herinneren van informatie, kunnen zowel KCR als EF bevorderlijk zijn voor de leerresultaten. KCR is doorgaans niet effectief als het om meer complexe denkniveaus gaat, terwijl EF dan juist wel voor betere leerresultaten zorgt (van der Kleij et al., 2011). Het huidige onderzoek richt zich op eenvoudige denkniveaus, wat het interessant maakt om de leereffecten van zowel de feedbackvormen KCR als EF te meten. De leereffecten bij beide feedbackvormen zijn gecontrasteerd met de feedbackvorm KR, die bij de controlegroep is aangeboden. Vanuit eerder onderzoek blijkt namelijk een beperkte positieve invloed van KR feedback (van der Kleij et al., 2015). Geen van deze drie feedbackvormen is aangevuld met *try again*, om te voorkomen dat het effect hiervan invloed heeft op de resultaten.

Op het gebied van de timing van feedback zijn positieve resultaten voor zowel directe als uitgestelde feedback gevonden, afhankelijk van het leerdoel van de les (Hattie & Timperley, 2007). Bij ingewikkelde taken wordt idealiter gebruik gemaakt van directe feedback, terwijl bij simpele taken uitgestelde feedback beter zou zijn voor de leerresultaten. Wanneer het gaat om het onthouden van procedurele of conceptuele kennis is directe feedback meer effectief. Om voor een transfer van de kennis te zorgen zou uitgestelde feedback de voorkeur verdienen (Shute, 2008). De focus van de serious game in dit onderzoek ligt op het onthouden van kennis, waardoor gebruik is gemaakt van directe feedback.

Bepaalde kenmerken zoals de schoolprestaties van de leerling, de achtergrond van de leertaak en de voorkennis hebben invloed op de effectiviteit van de verschillende feedbackvormen (Mason & Bruning, 2001; Narciss & Huth, 2004; Shute, 2008). Mason en Bruning (2001) hebben daarom een framework ontwikkeld voor het geven van feedback bij computergestuurde instructie waarin deze verschillende kenmerken zijn opgenomen. Bij basisschoolleerlingen kan worden verondersteld dat zij over het algemeen over een laag prestatieniveau en weinig voorkennis beschikken. Volgens het framework zouden zij daarom het meest gebaat zijn bij directe feedback in combinatie met KCR + EF *response contingent*.

Feedback kan alleen effectief zijn wanneer het op de juiste manier wordt aangeboden (Shute,

2008). Daarom wordt binnen dit onderzoek gebruik gemaakt van het stappenplan dat Narciss en Huth (2004) hebben samengesteld voor het ontwerpen van feedback. Zij beschrijven vijf stappen die moeten worden doorlopen om goede feedback te ontwerpen, namelijk: bepalen van de leerdoelen, leertaken selecteren, domein-specifieke kennis en vaardigheden vaststellen, veelgemaakte fouten en misconcepties achterhalen en op basis van deze kennis informatie specificeren die kan worden ingezet als feedback.

Instructiestrategieën

Naast feedback kan ook het gebruik van instructiestrategieën van invloed zijn op de leerresultaten. Zo beschrijven Roediger en Pyc (2012) drie instructie-strategieën die een positieve invloed hebben op de leerresultaten, namelijk: verspreiding van het leerwerk, tussentijds toetsen en het stellen van verklarende vragen. Bij het verspreiden van het leerwerk kan gebruik worden gemaakt van meerdere korte leersessies of het afwisselen van onderwerpen. Wanneer leerlingen het lesmateriaal in meerdere korte sessies bestuderen, blijken de leerresultaten op lange termijn beter te zijn dan wanneer zij alles in één keer leren (Rohrer & Pashler, 2010). Het afwisselen van verschillende onderwerpen waarbij hetzelfde onderwerp meerdere keren aan bod komt, blijkt op lange termijn ook effectiever te zijn dan wanneer ieder onderwerp in één aaneengesloten stuk wordt aangeboden (Roediger & Pyc, 2012). De tweede instructie-strategie richt zich op het tussentijds toetsen. McDaniel, Agarwal, Huelser, McDermott, en Roediger (2011) geven aan dat leerlingen beter presteren op de eindtoets wanneer zij tussentijdse toetsen krijgen. Een voorbeeld van een tussentijdse toets kan een laagdrempelige quiz zijn. Tot slot kan het stellen van verklarende vragen een positieve invloed hebben op de leerresultaten. Roediger en Pyc (2012) delen deze vragen op in twee soorten: vragen over de tekst en vragen over het leerwerk. Wanneer leerlingen zichzelf deze vragen stellen, kunnen zij de leerstof beter onthouden.

Volgens Dunlosky, Rawson, Marsh, Nathan, en Willingham (2013) heeft de tweede instructiestrategie, tussentijds toetsen, naar verwachting het meeste effect voor leerlingen in het basisonderwijs. Voor de andere twee instructiestrategieën, verspreiding van het leerwerk en het stellen van verklarende vragen, is nog te weinig bewijs gevonden bij jonge leerlingen. Daarom staat de instructiestrategie tussentijds toetsen centraal binnen dit onderzoek.

Ontwerp-framework voor serious games

Instructiestrategieën kunnen op allerlei manieren worden toegepast, zo ook in serious games. Deze digitale games zijn niet ontworpen voor commerciële doeleinden, maar voor het aanleren van bepaalde kennis of vaardigheden (Annetta, 2010). Serious games zijn gebaseerd op het principe dat leerlingen leren tijdens het spelen in een uitdagende omgeving waarin fouten mogen worden gemaakt (Papanastasiou et al., 2017). Het aanbod van serious games is de afgelopen jaren enorm toegenomen, maar niet alle serious games zijn even geschikt voor het onderwijs. Zo zijn serious games zelden

gebaseerd op achterliggende leertheorieën of wetenschappelijk onderzoek (Harteveld, Guimarães, Mayer, & Bidarra, 2010). Annetta (2010) heeft daarom een framework ontwikkeld dat bestaat uit een hiërarchisch model met zes elementen die samen het kader voor een educatieve game vormen. Om de effectiviteit van verschillende feedbackvormen te kunnen meten, moest de serious game binnen dit onderzoek (Plantendetective) van voldoende kwaliteit zijn. Daarom is het framework van Annetta gebruikt voor het ontwerp van deze serious game (zie Bijlage 1).

Het eerste element uit de hiërarchie, en hiermee de basis voor een serious game, is *identity*. Dit betreft het gevoel van de speler dat hij een uniek individu is binnen het spel. Het tweede element is *immersion*: onderdompeling in de game. Bij een goede game zijn spelers betrokken en intrinsiek gemotiveerd om het einddoel te behalen. Wanneer dit het geval is, kunnen zij in een staat van *flow* komen. Dit is een zeer actieve staat van concentratie en focus. Om deze staat te bereiken, zijn een duidelijk doel en geschikte feedback van groot belang (Kiili, 2005). Het volgende element is *interactivity*. Spelers moeten kunnen communiceren met andere spelers in het spel of met een computergestuurd medium. Vervolgens moet de complexiteit van het spel oplopen (*increasing complexity*), bijvoorbeeld door het gebruik van meerdere levels. Het vijfde element betreft *informed teaching*, dit gaat over de verwerking van feedback en beoordeling in het spel. Tot slot moet een serious game instructie bieden (*instructional*) om succesvol te zijn. De zone van naaste ontwikkeling staat hierbij centraal, leerlingen worden door de taken begeleid tot ze het zelfstandig kunnen. Binnen de game Plantendetective is dit gerealiseerd door de ondersteuning bij ieder nieuw level verder af te bouwen.

Met deze zes elementen is het mogelijk een goede serious game te ontwikkelen die zich op verschillende vakgebieden kan richten. In het voortgezet onderwijs zijn onderzoekers met name enthousiast over de inzet van serious games bij talenonderwijs (Wouters et al., 2013; Young et al., 2012), wiskunde (Divjak & Tomić, 2011), geschiedenis en bewegingsonderwijs (Young et al., 2012). Voor zowel het voortgezet- als het basisonderwijs neemt de laatste jaren de interesse in serious games binnen het vak natuuronderwijs toe (Barab & Dede, 2007; Li & Tsai, 2013; Wilson et al., 2018). Leerlingen zouden natuurkundige kennis en vaardigheden bij voorkeur in een rijke leeromgeving moeten ontwikkelen die verder gaat dan de muren van het klaslokaal. Een serious game biedt deze mogelijkheid (Barab & Dede, 2007). De game Plantendetective, die speciaal voor dit onderzoek is ontwikkeld, valt binnen het vak natuuronderwijs.

Serious games richten zich steeds vaker op mobiele platforms waardoor het mogelijk is om te leren buiten het klaslokaal (Giannakas, Kambourakis, Papasalouros, & Gritzalis, 2018). Zo kan een connectie worden gemaakt tussen de game zelf en de echte wereld, waardoor leerlingen hun omgeving op een andere manier gaan bekijken en actief leren gestimuleerd wordt (Huizenga, Admiraal, Akkerman, & ten Dam, 2009). Deze connectie kan bijvoorbeeld worden gemaakt door middel van AR

of door leerlingen zowel in het spel als in de echte wereld te laten oefenen met bepaalde handelingen (Li & Tsai, 2013). Door het gebruik van een camera en GPS is het mogelijk om het gedrag van de leerling in de echte wereld te volgen en hierbij persoonlijke begeleiding te bieden in de digitale omgeving (Giannakas et al., 2018). Gezien de mogelijkheden van mobiele games is de game Plantendetective met Seppo ontwikkeld en uitgeleverd. Dit platform maakte het mogelijk dat de leerlingen het spel via hun tablet in de natuurtuin van de school speelden, waar zij in aanraking kwamen met echte planten. Hierdoor konden zij tijdens de opdrachten zowel gebruik maken van de virtuele omgeving als de echte wereld.

De game Plantendetective is ontwikkeld aan de hand van de zes elementen van Annetta (2010). Om de kwaliteit van deze serious game te kunnen waarborgen, is Plantendetective voor de start van het interventieonderzoek technisch getest en geëvalueerd door een pilotgroep via een *cognitive walkthrough* en de afname van een kwaliteitsmeting. Tijdens deze *cognitive walkthrough* hebben de leerlingen in de pilotgroep de gehele serious game doorlopen en konden zij inzicht bieden in de manier waarop spelers denken en handelen terwijl zij Plantendetective spelen. Op basis van deze informatie en de resultaten uit de kwaliteitsmeting is Plantendetective vervolgens aangepast voor gebruik in het interventieonderzoek. Om de kwaliteit van Plantendetective te kunnen garanderen, is de game bovendien na afloop van het onderzoek door middel van een kwaliteitsmeting getoetst op de criteria van Annetta (2010).

Feedback in serious gaming

Serious games zijn steeds vaker digitaal waardoor gebruik kan worden gemaakt van computergestuurde feedback. Dit heeft een aantal belangrijke voordelen ten opzichte van menselijke feedback. Zo is computergestuurde feedback altijd onpartijdig, onbevooroordeeld en accuraat (Mason & Bruning, 2001). Daarnaast biedt een digitale serious game goede mogelijkheden voor directe feedback, de feedback kan immers al binnen een paar seconden na de reactie van de leerling worden gegeven (van der Kleij et al., 2015). Hierdoor kunnen leerlingen incorrecte informatie meteen aanpassen (Wouters et al., 2013). Tot slot nemen leerlingen feedback meer serieus wanneer ze het idee hebben dat het door een geloofwaardige bron wordt gegeven. Over het algemeen wordt computergestuurde feedback daarom beter ontvangen dan menselijke feedback (Shute, 2008). Anderzijds is het gemakkelijker om computergestuurde feedback te negeren (van der Kleij et al., 2015). Daarom is het bij het ontwerpen van feedback voor serious games belangrijk rekening te houden met de lengte en complexiteit van de informatie. Wanneer de tekst te lang en ingewikkeld is, besteden leerlingen nauwelijks aandacht aan de feedback (Shute, 2008). Daarnaast moet de feedback specifiek en duidelijk zijn en pas getoond worden nadat de leerlingen een antwoord hebben gegeven (Shute, 2008). Tot slot is het belangrijk dat feedback niet afleidt van het spel (Dunbar et al., 2017). Bij

Plantendetective is daarom voor elke vraag korte feedback ontworpen die specifiek in gaat op het onderwerp van de vraag. Deze feedback wordt getoond zodra de leerling zijn antwoord heeft ingestuurd en verdwijnt weer uit beeld wanneer de leerling door gaat naar de volgende vraag.

Vraagstellingen en hypothesen

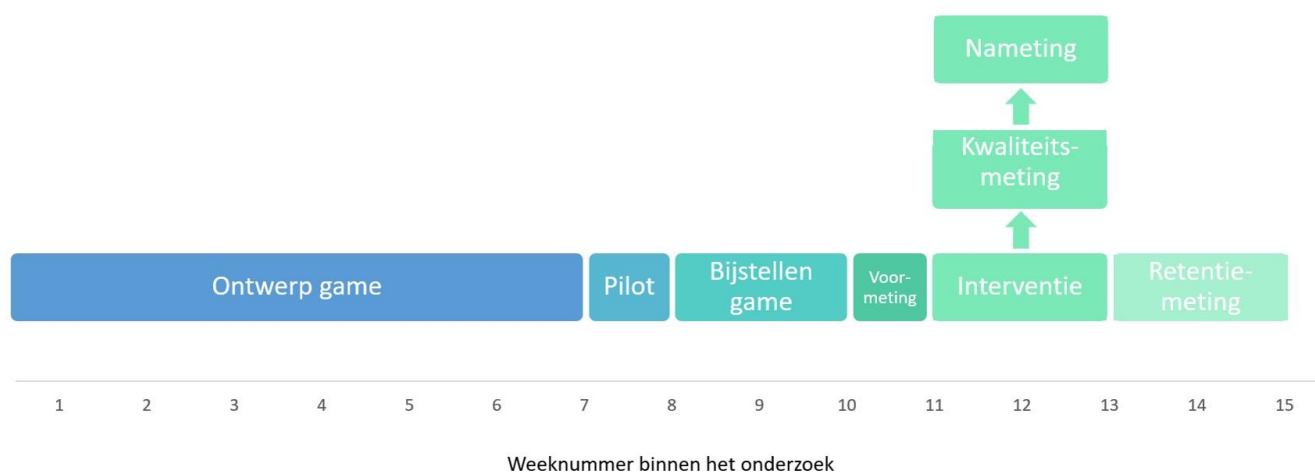
Dit onderzoek gaat na wat het leereffect is van verschillende feedbackvormen binnen serious gaming in het basisonderwijs. Hierbij worden de feedbackvormen EF en KCR, in twee experimentele groepen, en de feedbackvorm KR (controlegroep) met elkaar vergeleken binnen de door de onderzoeker ontworpen serious game Plantendetective. De onderzoeksvraag die hierbij centraal staat is: “Hoe effectief zijn de verschillende feedbackvormen voor het verbeteren van de leerresultaten van basisschoolleerlingen bij het spelen van een serious game over natuuronderwijs?” Vooraf is de verwachting dat EF en KCR beide effectiever zijn dan KR (Jaehnig & Miller, 2007). Daarnaast is de verwachting dat EF effectiever is dan KCR (van der Kleij et al., 2015). Naar aanleiding hiervan zijn de volgende hypothesen opgesteld:

- Hypothese 1: De leerresultaten van leerlingen in de KCR groep zijn beter dan de leerresultaten van leerlingen in de controlegroep.
- Hypothese 2: De leerresultaten van leerlingen in de EF *response contingent* groep zijn beter dan de leerresultaten van leerlingen in de controlegroep.
- Hypothese 3: De leerresultaten van leerlingen in de EF *response contingent* groep zijn beter dan de leerresultaten van leerlingen in de KCR groep.

Om deze hypothesen te kunnen testen, is het van belang dat de game Plantendetective van voldoende kwaliteit is. Daarom wordt deze op basis van het framework van Annetta (2010) ontworpen en ontwikkelde serious game vóór gebruik in een pilot geëvalueerd met een aantal leerlingen uit de doelgroep, via een *cognitive walkthrough*. Daarnaast wordt de kwaliteit van de game Plantendetective na afloop geëvalueerd aan de hand van een kwaliteitsmeting waarbij de zes elementen van Annetta (2010) worden getoetst. Hierbij is vooraf de verwachting dat de leerlingen het gevoel hebben onderdeel te zijn van het spel (*identity*), gemotiveerd zijn en in een *state of flow* kunnen komen (*immersion*), over voldoende mogelijkheden voor communicatie beschikken (*interactivity*), een goede balans tussen plezier en uitdaging ervaren (*increasing complexity*), voldoende feedback en toetsing krijgen (*informed teaching*) en genoeg informatie uit de game kunnen halen om de lesstof goed te begrijpen (*instructional*).

In Figuur 1 is het verloop van dit onderzoek aangegeven, per week is te zien welke activiteiten plaatsvinden. Voor de start van het interventieonderzoek wordt de game Plantendetective ontworpen

en gerealiseerd en vindt een pilot plaats op basis waarvan een definitieve versie van de game wordt gemaakt. Deze definitieve versie van de game wordt vervolgens gebruikt tijdens het interventieonderzoek.



Figuur 1. Volgordelijk verloop pilot en interventieonderzoek per week

Pilot

Voorafgaand aan het interventieonderzoek werd de game Plantendetective getest door een pilotgroep van drie leerlingen waarbij de bruikbaarheid en de inhoud van de game geëvalueerd werd door middel van een *cognitive walkthrough*. De leerlingen in de pilotgroep werden willekeurig geselecteerd uit alle participanten en namen vervolgens niet meer deel aan het interventieonderzoek. Zij mochten de serious game spelen en evalueren tijdens de reguliere lestijd, waardoor zij hier geen extra tijd mee kwijt waren. De leerlingen ervoeren dit als een eer, waardoor geen aanvullende beloning nodig was. Tijdens de *cognitive walkthrough* speelden de leerlingen Plantendetective terwijl zij hardop dachten en aan de onderzoeker vertelden wat zij van het spel en de opdrachten vonden. Hieruit bleek vooral dat de besturing van de game niet altijd even duidelijk was. Bij de vragen waarbij de leerlingen in de pilotgroep niet precies wisten waar ze op moesten klikken of wat ze moesten doen, is daarom voor de definitieve versie van Plantendetective extra toegelicht op welke knoppen de leerlingen moeten klikken (zie Figuur 2). Daarnaast waren twee van de drie leerlingen niet bekend met de varen. Daarom is besloten om wat meer informatie te geven over deze plant (zie Figuur 3).

Om te kunnen onderzoeken of de game Plantendetective aan de kwaliteitscriteria van Annetta (2010) voldoet, vulden de leerlingen in de pilotgroep na afloop een kwaliteitsmeting in (zie Bijlage 3) waarmee de kwaliteit van de onderdelen *identity*, *immersion*, *complexity*, *interactivity*, *informed*

teaching en *instruction* werd gemeten. In de methodesectie wordt onder het kopje meetinstrumenten verder toegelicht hoe deze kwaliteitsmeting tot stand is gekomen.

De leerlingen uit de pilotgroep vulden de kwaliteitsmeting na het spelen van de game Plantendetective in terwijl zij mondeling aan de onderzoeker toelichtten waarom ze bepaalde scores toekenden. Hierbij viel vooral het onderdeel *interactivity* op, dat onvoldoende scoorde. Uit het gesprek met de leerlingen bleek dat zij zich niet goed bewust waren van de mogelijkheid om via de chatbox vragen te stellen over de game. Daarom werd besloten een extra uitleg toe te voegen (zie Figuur 4) om de leerlingen op deze mogelijkheid te wijzen. De onderdelen *identity*, *immersion*, *complexity*, *informed teaching* en *instruction* scoorden allemaal een ruime voldoende. Op basis hiervan werd besloten geen verdere wijzigingen door te voeren in de definitieve game.



Figuur 2. Extra toelichting bij de besturing van het spel



Figuur 3. Extra uitleg over varens



Figuur 4. Extra uitleg in de game over het gebruik van de chatbox, die rechtsonder in het hoofdscherm zichtbaar is. Via de knop ‘voorvertoning’ kunnen de leerling hun antwoord bekijken. Door op ‘verstuur’ te klikken, sturen zij het antwoord in

Methode

Ontwerp

In dit interventieonderzoek werd een gerandomiseerd pretest-posttest-retentietest ontwerp naar de leereffecten van feedbackvormen gevolgd, met twee experimentele groepen (KCR en EF *response contingent*) en een controlegroep (KR). Gezien het feit dat het ethisch niet verantwoord is een groep leerlingen geen enkele vorm van feedback te geven, kregen de leerlingen in de controlegroep de feedbackvorm KR aangeboden. Eerder onderzoek toont namelijk een beperkter positief leereffect van KR in vergelijking met andere feedbackvormen (van der Kleij et al., 2015).

De leerlingen in dit onderzoek werden willekeurig over de drie onderzoeksgroepen verdeeld. Alle leerlingen kregen drie metingen: een voormeting, een nameting en een retentiemeting. Bij elke meting werd de kennis op het gebied van natuuronderwijs gemeten, dit is de afhankelijke variabele. Na de voormeting speelden de leerlingen een serious game. Hierbij werd beoogd de kennistoename te manipuleren door middel van feedback. De feedbackvorm die de leerlingen kregen, is de onafhankelijke variabele.

Participanten

Voor dit onderzoek werden de leerlingen van een grote basisschool in Zuid-Holland geselecteerd om deel te nemen. Aangezien alleen de bovenbouwgroepen les in natuuronderwijs krijgen, werd aan de leerlingen van de drie jaargroepen 7 en drie jaargroepen 8 gevraagd te participeren. Leerlingen uit groep 7 of 8 die via een eigen leerlijn op een lager niveau les krijgen, werden uitgesloten van deelname aan dit onderzoek. Van de 175 leerlingen in de groepen 7 en 8 werden uiteindelijk vier leerlingen uitgesloten vanwege dit exclusie criterium. Hierdoor komt de totale onderzoeksgroep uit op 171 leerlingen. Dit aantal is groter dan het minimaal benodigde aantal van 141 participanten dat gevonden is op basis van een berekening in G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) (zie Bijlage 2). Voor deze berekening is uitgegaan van een effectgrootte van Cohen's $d = .30$. Deze effectgrootte is kleiner dan de effectgrootte van Cohen's $d = .49$ die van der Kleij et al. (2015) in hun meta-analyse vonden voor de feedbackvorm EF. Zij hebben echter ook meer uitgebreide vormen van EF onderzocht, terwijl in dit onderzoek wordt gefocust op *response contingent* EF, een specifieke en vrij compacte vorm van EF. Daarom is bij de berekening van het minimaal benodigde aantal participanten in dit onderzoek gebruik gemaakt van een lagere effectgrootte.

Aan het feit dat alle participanten bij elkaar op school zitten, is een belangrijk nadeel verbonden. Het was op deze manier namelijk onmogelijk om de leerlingen uit verschillende onderzoeksgroepen van elkaar te scheiden, waardoor zij hun ervaringen binnen het onderzoek met elkaar zouden kunnen delen. Deze *diffusion of treatments* is een bedreiging voor de interne validiteit. Om deze bedreiging te beperken, vond de eerste nameting op dezelfde dag plaats als de interventie.

Een voordeel van het feit dat alle participanten bij elkaar op school zitten, is dat de invloed van externe factoren, zoals geografische ligging en opleidingsniveau van ouders, zo beperkt mogelijk blijft en de praktische haalbaarheid van het onderzoek wordt vergroot. Daarom is toch de keuze gemaakt met slechts één basisschool te werken.

Een lage startrespons en uitval van leerlingen vormden eveneens een bedreiging van de validiteit en de betrouwbaarheid van de resultaten, bijvoorbeeld wanneer ouders geen toestemming zouden geven voor deelname aan het onderzoek of wanneer leerlingen niet aanwezig zouden kunnen zijn bij één van de drie meetmomenten. Hierdoor konden onder andere *missing values* optreden die de statische power van het onderzoek verkleinen. Om de startrespons zo hoog mogelijk te maken, is proactief contact gezocht met ouders die niet reageerden op de toestemmingsbrief van dit onderzoek. Hen werd herhaaldelijk gevraagd om dit alsnog te doen. Aangezien de meetmomenten binnen de reguliere schooltijden vielen, was de verwachting dat de leerlingen in principe bij alle drie de meetmomenten aanwezig zouden zijn. Hierdoor zijn geen aanvullende maatregelen getroffen om uitval te voorkomen. Uiteindelijk zijn twee leerlingen niet bij alle meetmomenten aanwezig geweest. Hun resultaten konden daarom niet worden meegenomen in dit onderzoek (N = 140).

Materialen

Speciaal voor dit onderzoek ontwierp de onderzoeker een serious game over het vak natuuronderwijs. Deze serious game, Plantendetective, kan buiten het klaslokaal in maximaal 45 minuten worden uitgespeeld. Plantendetective werd ontwikkeld met behulp van het platform Seppo en richt zich op het onderwerp ‘het plantenrijk’. Zowel de leerlingen uit jaargroep 7 als jaargroep 8 hadden hier nog nooit eerder les over gehad. Het spel bestaat uit drie levels in oplopende moeilijkheidsgraad. Elk level bestaat uit instructie in de vorm van tekst en afbeeldingen waarbij de leerling informatie krijgt over het onderwerp. Vervolgens moet de leerling vier meerkeuzevragen beantwoorden om door te kunnen gaan naar het volgende level. Deze meerkeuzevragen vervullen de rol van een tussentoets waarbij de leerlingen hun opgedane kennis kunnen meten. Een voorbeeld van een meerkeuzevraag is: *Wat voor soort plant is dit? a.) een zaadplant b.) een sporenplant*. Meteen na het indienen van het antwoord kreeg de leerling feedback te zien, afhankelijk van de groep waar hij in zat betrof dit KR, KCR of EF *response contingent* feedback. De KR groep kreeg een ‘goed’ of ‘fout’ melding, bij KCR kreeg de leerling naast ‘fout’ ook te zien wat wel het goede antwoord zou zijn geweest. In de EF *response contingent* feedbackgroep kreeg de leerling bij een correct antwoord te zien dat het antwoord goed is, hierbij werd uitgelegd waarom dit zo is. Als de leerling het antwoord fout had, kreeg hij het goede antwoord te zien gecombineerd met de uitleg waarom dit antwoord goed is.

Alle leerlingen konden met een correct antwoord 50 punten verdienen. Deze punten werden tegelijk met de feedback aan de leerlingen getoond. Gedurende het spel konden zij bovendien hun

eigen voortgang bijhouden en continu zien hoeveel punten zij in totaal hadden gescoord. Om Plantendetective te kunnen spelen, hadden alle leerlingen de beschikking over een tablet die door de school werd verstrekt. Met deze tablet konden zij inloggen via de website (www.seppo.io). Hiervoor kregen de leerlingen een geanonimiseerd account dat gekoppeld was aan een code. Met deze code konden zij altijd opnieuw inloggen en weer verder gaan waar zij gebleven waren.

Meetinstrumenten

Tijdens het interventieonderzoek werd gebruik gemaakt van kennistesten om de voorkennis en de kennistoename bij de leerlingen te kunnen meten. Aangezien geen gevalideerde testen beschikbaar waren die de kennis over het plantenrijk meten, ontwikkelde de onderzoeker zelf drie kennistesten die aansluiten op de lesstof over het plantenrijk voor de voor-, na- en retentiemeting (zie Bijlage 4). Om voor voldoende validiteit te zorgen, zijn de kennistesten beoordeeld door twee experts op het gebied van natuuronderwijs. De inhoud van de kennistesten sluit aan op de leerdoelen en inhoud van Argus Clou, een lesmethode voor natuur en techniek in het basisonderwijs. Elke kennistest werd afgenomen op papier en bestond uit tien meerkeuze vragen, deze vragen zijn namelijk gemakkelijk te beoordelen zonder dat de interpretatie van de onderzoeker hierbij een rol speelt. Een voorbeeld van een meerkeuzevraag uit de kennistesten is: *Waaruit bestaat het plantenrijk? a. Sporenplanten en schimmels b. Schimmels en zaadplanten c. Sporenplanten en zaadplanten*. Om de kennistesten zo goed mogelijk met elkaar te kunnen vergelijken, werden steeds drie vragen tegelijk opgesteld (één voor iedere kennistest). Deze vragen werden zo gemaakt dat de opzet van de vraag steeds hetzelfde is, maar de formulering en antwoordmogelijkheden van elkaar verschillen. Op deze manier zijn de vragen niet exact hetzelfde, waardoor het risico op *testing effects* zo beperkt mogelijk blijft. De kennistesten zijn echter wel goed met elkaar te vergelijken. Om uit te sluiten dat het verschil tussen de drie kennistesten invloed zou hebben op de resultaten, bijvoorbeeld doordat één van de kennistesten gemakkelijker zou zijn dan de andere twee, werden de kennistesten door elkaar heen gebruikt. Dat wil zeggen dat alle drie de kennistesten tijdens de eerste meting werden gebruikt waarbij zij gelijk werden verdeeld over de onderzoeksgroepen.

Naast de kennistesten kon ook de Seppo omgeving worden gebruikt om informatie over de participanten te verzamelen. Via deze omgeving werd bijvoorbeeld geregistreerd of de leerlingen de game Plantendetective volledig hadden afgerond, hoe lang zij bezig waren geweest met de game (sessieduur) en hoeveel punten zij in totaal tijdens de game gehaald hadden (gamescore).

Om de kwaliteit van de game Plantendetective te meten, werd gebruik gemaakt van een kwaliteitsmeting waarmee de zes elementen van Annetta (2010) konden worden getoetst (zie Bijlage 3). Deze kwaliteitsmeting werd samengesteld uit onderdelen van de Flow-Short (Rheinberg et al., 2003) voor het meten van de flow van de game, de NASA-TLX (Hart & Staveland, 1988) voor het

meten van de taakbelasting, de Intrinsic Motivation Inventory (Ryan & Deci, 2000) voor het meten van de motivatie van leerlingen, een set vragen afkomstig uit het onderzoek van Nadolski en Hummel (2017) bedoeld voor het meten van de studeerbaarheid en een aantal aanvullende vragen. Om de zes elementen van Annetta (2010) te kunnen toetsen, werden de scores die de leerlingen op de verschillende onderdelen van de kwaliteitsmeting toekenden, omgezet in gestandaardiseerde scores. Dit was nodig omdat voor sommige vragen een vijfpuntsschaal werd toegepast, terwijl voor andere vragen een zevenpuntsschaal werd gehanteerd. Daarom is de keuze gemaakt alle scores om te vormen naar een gestandaardiseerde score tussen 0 en 1, waarbij een score hoger dan .5 als voldoende werd beschouwd. Met behulp van deze gestandaardiseerde score kon een totaalscore worden berekend voor de zes elementen van Annetta (2010). Hierbij is gebruik gemaakt van een gewogen gemiddelde aan de hand van het aantal items per subonderdeel. Voor de vragen die negatief gesteld werden, is de score omgekeerd ingevoerd in SPSS. De vragen waarvoor dit geldt zijn in Bijlage 3 aangegeven met een sterretje.

4.4 Procedure

Voor de start van de pilot en het interventieonderzoek werd contact opgenomen met de geselecteerde basisschool. Door middel van een schriftelijk ondertekend *informed consent* formulier konden de directie, de leerlingen en de ouders van de leerlingen toestemming geven voor deelname aan het onderzoek. Vervolgens werd een aantal persoonsgegevens van de deelnemers verzameld, namelijk het geslacht, de leeftijd, jaargroep en het niveau op basis van de Citoresultaten en het schooladvies. Hierbij werd het anonimiseren van de gegevens expliciet gemeld. Alle leerlingen werden op basis van de namenlijst over de drie onderzoeksgroepen verdeeld. Hierbij werd de eerste leerling op de lijst in de eerste groep geplaatst, de tweede leerling in de tweede groep enzovoort. Deze randomisatie controleerde de invloed van externe variabelen, waaronder prestatieniveau en voorkennis over het onderwerp van de les (Mason & Bruning, 2001). De leerlingen uit jaargroep 7 / 8 werden hierbij gelijk over de groepen verdeeld, omdat vooraf verwacht werd dat leerlingen uit jaargroep 8 een hoger prestatieniveau en meer voorkennis zouden hebben dan leerlingen uit jaargroep 7. Deze gelijke verdeling werd gerealiseerd door eerst alle leerlingen van jaargroep 7 willekeurig over de onderzoeksgroepen te verdelen, gevolgd door de leerlingen uit jaargroep 8. Dit betekent dat de onderzoeksgroepen gevormd werden door leerlingen die onder normale omstandigheden niet bij elkaar in de klas zitten. Alle leerlingen kregen vervolgens een code toegekend. Welke leerling bij welk code hoort, is alleen bij de onderzoeker bekend. De privacy van de leerlingen wordt hiermee gewaarborgd.

Vervolgens kregen alle leerlingen een voormeting die werd afgenomen in de eerste week van het experiment. De leerlingen maakten de kennistest van de voormeting individueel in hun eigen lokaal in het bijzijn van de groepsleerkracht en de onderzoeker. In de tweede en derde week van het

experiment speelden de leerlingen de game Plantendetective. Alle leerlingen waren via de game omgeving gekoppeld aan hun onderzoeksgroep en kregen, afhankelijk van de groep waarin zij ingedeeld waren, KR, KCR of EF als feedback aangeboden. Om het risico van het doorgeven van informatie tussen de verschillende groepen zo klein mogelijk te houden, speelden alle leerlingen uit de jaargroepen 7 op dezelfde dag de game en waren de week daarna alle leerlingen uit de jaargroepen 8 aan de beurt. Plantendetective werd gespeeld in de natuurtuin naast de school onder toezicht van de onderzoeker. De leerlingen speelden de game individueel en waren hier in totaal maximaal 45 minuten mee bezig. Om Plantendetective te kunnen spelen, kregen alle leerlingen een tablet mee van de school. Met de code die aan het begin van het onderzoek aan de leerlingen was toegekend, konden zij inloggen in de spelomgeving. De onderzoeker vulde deze code voor de leerling in bij de start van de game. Om te voorkomen dat de leerlingen met elkaar konden overleggen, begonnen zij één voor één aan de serious game. Daarnaast hield de onderzoeker toezicht en voorkwam dat leerlingen toch contact met elkaar zochten. Leerlingen uit verschillende onderzoeksgroepen konden Plantendetective gelijktijdig spelen. Door hun inlogcode waren zij namelijk automatisch aan de juiste vorm van feedback gekoppeld.

Alle leerlingen speelden Plantendetective op maandagochtend en vulden direct na de game de kwaliteitsmeting in. Vervolgens vulden zij in de middag de kennistest van de nameting in. De nameting werd evenals de voormeting in het eigen lokaal afgenomen in het bijzijn van de groepsleerkracht en de onderzoeker. Twee weken later kwam de onderzoeker terug in de klas om een retentiemeting af te nemen, dit gebeurde op dezelfde wijze als de voor- en nameting en de leerlingen vulden de kennistest van deze retentietest eveneens in de middag in.

4.5 Data-analyse

De data-analyses die gebruikt werden om de hypothesen te toetsen, werden uitgevoerd in SPSS versie 26. In het ontwerp van dit onderzoek is één between-subject variabele opgenomen (feedbackvorm) en één within-subject variabele (verschilsscores op de kennistoets bij de voor-, na- en retentiemeting op het gebied van plantenkennis). Het effect van deze twee onafhankelijke variabelen op de afhankelijke variabele (kennistoename) werd getoetst aan de hand van een mixed ANOVA.

De resultaten van de mixed ANOVA werden getoetst tegen een significantieniveau van 0.05. Op aanraden van van der Kleij et al. (2015) werd de referentiemaat van Hattie (2009) gebruikt voor het interpreteren van de effectgrootte. Deze referentiemaat is namelijk gevalideerd voor gebruik in het onderwijs, in tegenstelling tot de referentiemaat van Cohen (1988), die gewoonlijk wordt gebruikt voor het interpreteren van de effectgrootte. Volgens Hattie (2009) kan een effectgrootte van 0.2 als klein worden beschouwd, een effectgrootte van 0.4 kan als gemiddeld worden gezien en een effectgrootte van 0.6 kan als groot worden geclassificeerd.

De kwaliteit van de game Plantendetective werd geëvalueerd door gestandaardiseerde scores te berekenen (tussen 0 en 1) op de zes elementen van Annetta (2010) die onderdeel zijn van de kwaliteitsmeting die leerlingen na afloop van de game invulden. Bij een gemiddeld meer positieve dan negatieve score (wat het geval is bij een score hoger dan .5) werd een onderdeel als voldoende beoordeeld.

Resultaten

Van de 171 leerlingen kregen uiteindelijk 142 leerlingen toestemming voor deelname aan het onderzoek. Hiervan konden twee leerlingen het onderzoek niet volledig afronden. Dit resulteert in een totaal aantal van 140 participanten ($N = 140$) verdeeld over drie verschillende onderzoeksgroepen: KR, KCR en EF. In Tabel 1 zijn de demografische gegevens van deze participanten per onderzoeksgroep gespecificeerd. Hierin is te zien dat in het totale onderzoek meer jongens deelnamen dan meisjes. De verhouding van jongens en meisjes blijkt niet significant te verschillen over de experimentele groepen [$\chi^2(2) = 2.156, p = .34$], waardoor kan worden gesteld dat er sprake is van een gelijke verdeling. Verder blijkt ook de gemiddelde leeftijd bij de verschillende onderzoeksgroepen niet significant van elkaar te verschillen [$F(2.137) = .187, p = .830$]. Tot slot zijn de niveaus van de leerlingen gegroepeerd in drie groepen: laag, gemiddeld en hoog. De lage groep bevat alle leerlingen met een Vmbo Basis, Vmbo Basis / Kader of een Vmbo Kader advies voor de middelbare school. In de gemiddelde groep zitten de leerlingen met een Vmbo Kader / Theoretisch, Vmbo Theoretisch of Vmbo Theoretisch / Havo advies. Het hoge niveau omvat de Havo, Havo / Vwo en Vwo adviezen. Het grootste deel van de participanten behoort tot de groep met een hoog niveau, slechts een klein aantal participanten had een laag niveau. Hierbij zijn geen significante niveauverschillen zichtbaar tussen de drie experimentele groepen [$\chi^2(4) = 2.652, p = .618$], waardoor kan worden verondersteld dat het niveau van de leerlingen gelijk over de groepen is verdeeld. Vooraf was de verwachting dat leerlingen uit groep 8 meer voorkennis en dus een hogere score op de voormeting zouden halen dan leerlingen uit groep 7, maar dit blijkt niet het geval te zijn. De gemiddelde score van de leerlingen uit groep 8 ($M = 5.25, SD = 1.992$) is weliswaar hoger dan die van de leerlingen uit groep 7 ($M = 4.86, SD = 1.572$), maar een t-toets laat zien dat dit verschil niet significant is, $t(138) = -1.31, p = .192$. Ook voor de scores op de nameting en retentiemeting konden geen significante verschillen worden gevonden tussen de leerlingen uit groep 7 en groep 8.

Tabel 1

Demografische gegevens per onderzoeksgroep

	N	Geslacht		Jaargroep		Leeftijd	Niveau		
		J	M	7	8	M (SD)	Laag	Gemiddeld	Hoog
KR	45	23	22	26	19	138 (7.6)	7	13	25
KCR	48	27	21	27	21	138 (8.7)	6	18	24
EF	47	31	16	23	24	137 (8.2)	3	18	26

Noot. Gemiddelde leeftijd is gespecificeerd in maanden.

Leereffecten van de game

Voor en na het spelen van Plantendetective werden kennistesten afgenomen om de planten kennis van de leerlingen te meten. De resultaten van deze kennistesten worden beschreven in Tabel 2 en grafisch weergegeven in Figuur 5. Mauchly's test voor sphericiteit laat zien dat de variantie van de vershilscores op deze testen tussen de onderzoeksgroepen gelijk is, $X(2) = 1.127, p = .569$. Hiermee wordt voldaan aan de assumptie van sphericiteit voor het uitvoeren van een mixed ANOVA. Uit de resultaten van deze mixed ANOVA blijkt dat de gemiddelde scores van alle leerlingen significant verschillen op de drie meetmomenten, $F(2,274) = 38.477, p < .001$. De scores op de nameting zijn significant hoger dan op de voormeting $F(1,137) = 79.384, p < .001$ en de scores op de retentiemeting zijn eveneens hoger dan op de voormeting, maar wel significant lager dan de nameting $F(1,137) = 19.511, p < .001$. Dit betekent dat de serious game Plantendetective een significant persistent leereffect veroorzaakt. De effectgrootte tussen de voor- en nameting is volgens de interpretatie van Hattie (2009) groot ($d = .81$). Het leereffect tussen de voor- en retentiemeting is klein ($d = .39$).

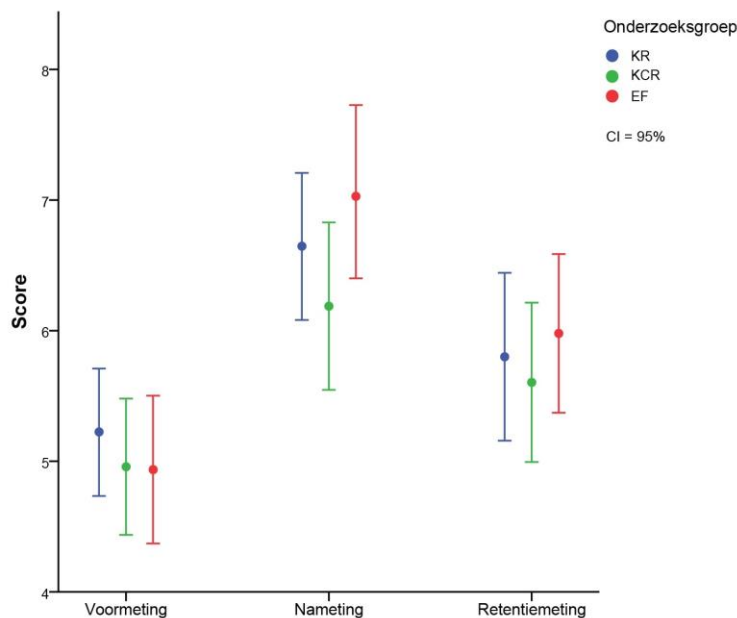
Het belangrijkste doel van dit onderzoek is het vergelijken van het effect van drie feedbackvormen op de leerresultaten van de serious game bij basisschoolleerlingen. Om die reden zijn de hypothesen in dit onderzoek zodanig geformuleerd dat deze drie feedbackvormen met elkaar worden vergeleken. Uit de resultaten van de mixed ANOVA blijkt echter dat er geen significant interactie effect is tussen de feedbackvorm en de scores op de verschillende meetmomenten $F(4,274) = 1.190, p = .316$. Dit betekent dat de feedbackvorm geen invloed heeft op het leereffect van de serious game en dat dus geen ondersteuning gevonden kan worden voor de hypothesen die in dit onderzoek zijn opgesteld.

Om de betrouwbaarheid van de zelfontwikkelde kennistesten te kunnen onderzoeken, is een betrouwbaarheidsmeting gedaan. Uit deze betrouwbaarheidsmeting blijkt dat de drie kennistesten laag- tot matig betrouwbaar zijn (kennistest 1: 10 items, $\alpha = .544$; kennistest 2: 10 items, $\alpha = .492$; kennistest 3: 10 items, $\alpha = .559$).

Tabel 2

Gemiddelde scores op de voor-, na- en retentiemeting per onderzoeksgroep met een spreiding van 95% CI

	Conditie	<i>M</i>	<i>SE</i>	CI (95%)
Voormeting [0-10]	KR	5.22	.27	[4.70 – 5.75]
	KCR	4.96	.26	[4.45 – 5.47]
	EF	4.94	.26	[4.42 – 5.45]
Nameting [0-10]	KR	6.64	.32	[6.01 – 7.27]
	KCR	6.19	.31	[5.58 – 6.79]
	EF	7.06	.31	[6.45 – 7.68]
Retentiemeting [0-10]	KR	5.80	.31	[5.18 – 6.42]
	KCR	5.60	.30	[5.00 – 6.20]
	EF	5.98	.31	[5.37 – 6.59]



Figuur 5. Grafische weergave van de gemiddelde scores op de voor-, na- en retentiemeting per onderzoeksgroep met een spreiding van 95% CI

Gamescores

Alle participanten speelden individueel de game Plantendetective waarmee zij een maximale gamescore van 840 punten konden behalen. In Tabel 3 is de gemiddelde gamescore van de onderzoeksgroepen KR, KCR en EF weergegeven. Hierin is te zien dat de participanten in de EF groep gemiddeld de hoogste gamescore behaalden. Daarmee verschilt de EF groep significant van de KR groep [$t(90) = -3.100, p = .003$] en van de KCR groep [$t(93) = -3.234, p = .002$]. De KR en KCR groep verschillen niet significant van elkaar [$t(91) = -0.207, p = .837$].

Sessieduur

In Tabel 3 is ook de gemiddelde sessieduur voor de game Plantendetective per onderzoeksgroep terug te vinden. Hieruit blijkt dat de gemiddelde sessieduur van de EF groep het hoogst is met ruim twee minuten meer dan de andere twee groepen. Voor de sessieduur zijn echter geen significante verschillen gevonden.

Tabel 3

Gamescore (aantal punten) en sessieduur (in minuten) voor het spelen van de serious game

	Conditie	M	SD	Min.	Max.
Aantal punten	KR	523	130	270	750
	KCR	518	141	230	780
	EF	609	133	250	840
Sessieduur	KR	16:07	06:45	05:24	37:38
	KCR	15:42	07:02	04:02	37:05
	EF	18:18	06:43	06:46	38:15

Kwaliteit game

Naast de nameting vulden de leerlingen een kwaliteitsmeting in met vragen voor het meten van de kwaliteit van de serious game (zie Bijlage 3). Met behulp van deze kwaliteitsmeting werd Plantendetective door de leerlingen geëvalueerd op de zes elementen van Annetta (2010). De scores die de leerlingen toekenden aan deze verschillende onderdelen zijn te vinden in Tabel 4. Hierbij is voor alle onderdelen een gestandaardiseerde score weergegeven. Het element *identity* werd gemeten aan de hand van twee zelfontwikkelde vragen. Dit element is in totaal uitgekomen op de gestandaardiseerde score van .625. *Immersion* werd gemeten aan de hand van de *Intrinsic Motivation Inventory* (IMI) van Ryan en Deci (2000), de Flow-short van Rheinberg et al. (2003) en studeerbaarheidsvragen van Nadolski en Hummel (2017). De gewogen totaalscore van deze onderdelen komt uit op .698. Het element *Interactivity* werd gemeten met behulp van drie zelfontwikkelde vragen. Hierbij werd een gestandaardiseerde score van .750 behaald. Voor het meten van de *increasing complexity* is gebruik gemaakt van een combinatie aan vragen uit de NASA-TLX (Harteveld et al., 2010) en studeerbaarheidsvragen (Nadolski & Hummel, 2017). Het gewogen gemiddelde hiervan komt uit op .750. *Informed teaching* en *instruction* werden eveneens getoetst aan de hand van studeerbaarheidsvragen, waarbij gestandaardiseerde scores van .650 en .675 werden behaald. Hiermee scoorden alle elementen van de kwaliteitsmeting bovengemiddeld (hoger dan .5), waarmee de kwaliteit van de game Plantendetective als voldoende kan worden beschouwd.

Voor de verschillende onderdelen van de kwaliteitsmeting werd achteraf een Cornbach's alpha (α) berekend voor het meten van de betrouwbaarheid. Uit deze betrouwbaarheidsanalyse blijkt dat de IMI een zeer hoge betrouwbaarheid heeft (11 items, $\alpha = .915$). De betrouwbaarheid van de

Studeerbaarheidsvragenlijst bleek eveneens erg hoog (14 items, $\alpha = .851$). Zowel de Flow-Short (10 items, $\alpha = .741$) als de aanvullende vragenlijst (5 items, $\alpha = .715$) bleek voldoende betrouwbaar te zijn. Alleen de NASA-TLX bleek een matige betrouwbaarheid te hebben (6 items, $\alpha = .541$).

Tabel 4

Resultaten kwaliteitsmeting per element van Annetta (2010)

Element	Onderdeel kwaliteitsmeting	Aantal items	Gestandaardiseerd gemiddelde
Identity	Zelf ontwikkeld	2	.625
Immersion	IMI	11	.683
	Flow-short	10	.700
	Studeerbaarheid	5	.725
Interactivity	Zelf ontwikkeld	3	.750
Increasing complexity	NASA-TLX	6	.767
	Studeerbaarheid	2	.700
Informed Teaching	Studeerbaarheid	4	.650
Instruction	Studeerbaarheid	3	.675

Conclusie en discussie

Dit onderzoek met een gerandomiseerd pretest-posttest-retentietest ontwerp had als doel meer kennis te verwerven over de effectiviteit van verschillende feedbackvormen binnen serious gaming in het basisonderwijs. Hierbij stond de vraag “Hoe effectief zijn de verschillende feedbackvormen voor het verbeteren van de leerresultaten van basisschoolleerlingen bij het spelen van een serious game over natuuronderwijs?” centraal. Om antwoord te kunnen geven op deze vraag, werden de feedbackvormen EF en KCR, in twee experimentele groepen, vergeleken met de feedbackvorm KR (controlegroep) binnen een door de onderzoeker ontwikkelde serious game. De verwachting was hierbij dat EF en KCR beide effectiever zouden zijn dan KR en dat EF effectiever zou zijn dan KCR. Vooraf werden op basis hiervan de volgende drie hypothesen opgesteld:

- Hypothese 1: De leerresultaten van leerlingen in de KCR groep zijn beter dan de leerresultaten van leerlingen in de controlegroep.
- Hypothese 2: De leerresultaten van leerlingen in de EF *response contingent* groep zijn beter dan de leerresultaten van leerlingen in de controlegroep.
- Hypothese 3: De leerresultaten van leerlingen in de EF *response contingent* groep zijn beter dan de leerresultaten van leerlingen in de KCR groep.

Uit de resultaten van de kennistesten voor de drie meetmomenten blijkt dat de gemiddelde score op de kennistest van de voormeting voor alle drie de onderzoeksgroepen nagenoeg gelijk is. Dit betekent dat de leerlingen in alle groepen ongeveer evenveel voorkennis hebben over het onderwerp van de game. Na het spelen van Plantendetective is bij alle groepen een leereffect zichtbaar, aangezien de leerlingen significant hoger scoren op de nameting dan op de voormeting. Na twee weken dooft dit leereffect gedeeltelijk uit, maar ondanks deze daling is nog steeds een significante verbetering voor alle onderzoeksgroepen zichtbaar tussen de voor- en retentiemeting. Op de kennistest van de nameting scoorden de leerlingen in de KCR groep niet hoger dan de leerlingen in de KR groep (hypothese 1). De KR groep haalde gemiddeld zelfs een iets hogere score, maar dit verschil is niet significant. De EF groep haalde gemiddelde de hoogste score op de kennistest van de nameting. Dit gemiddelde blijkt echter ook niet significant te verschillen van de KR groep (hypothese 2) of de KCR groep (hypothese 3). Tot slot blijkt ook uit de resultaten van de kennistesten van de retentiemeting dat de leerlingen in de EF groep gemiddeld de hoogste score halen, maar deze score verschilde wederom niet significant van de KR en KCR groep. Op basis van deze resultaten moeten alle drie de hypothesen worden verworpen, waardoor kan worden geconcludeerd dat de feedbackvorm geen invloed heeft op de leerresultaten van de serious game Plantendetective.

Uit de resultaten blijkt dat er geen significante verschillen werden gevonden in de tijd die de verschillende groepen aan de serious game besteedden (sessieduur). Voor de game-score blijkt echter dat de EF groep een significant hogere game-score haalde dan de KR en KCR groep. Deze hogere gamescore is opvallend, omdat de EF groep de voor-, na- en retentiemeting geen significant hogere scores haalde dan de andere twee groepen. Het is overigens de vraag of dit significante verschil verklaard kan worden door de aangeboden feedbackvorm, aangezien de feedback pas gegeven wordt nadat het aantal punten is toegekend. Het is dus niet duidelijk hoe dit verschil veroorzaakt wordt.

Aangezien Plantendetective een zichtbaar leereffect veroorzaakt, kan worden aangenomen dat de game van voldoende kwaliteit is. Deze aanname wordt ondersteund door de resultaten van de kwaliteitsmeting die de leerlingen na afloop van de game hebben ingevuld. Hieruit blijkt dat Plantendetective op alle zes de elementen van Annetta (2010) voldoende scoort. Dit is een verbetering ten opzichte van de resultaten van de eerste versie van de game die bij de pilot werd gebruikt, hierbij scoorde het onderdeel *interactivity* namelijk nog onvoldoende. Dit betekent dat de wijzigingen die naar aanleiding van de pilot werden toegepast, een positief effect hebben gehad op de kwaliteit van de game Plantendetective.

Vergelijking met eerder onderzoek

De resultaten van dit onderzoek komen niet overeen met de bevindingen uit eerder onderzoek naar de verschillen in leereffecten tussen KR, KCR en EF. Zo blijkt uit onderzoek van onder andere Jaehnig

en Miller (2007), Mason en Bruning (2001) en van der Kleij et al. (2015) dat EF tot significant betere leerresultaten kan leiden dan KR en KCR. Deze onderzoeken zijn echter niet in het basisonderwijs uitgevoerd. In hun meta-analyse hebben van der Kleij et al. (2015) een beperkt aantal resultaten uit het basisonderwijs meegenomen. Zij vonden op basis hiervan kleinere effectgroottes voor deze onderzoeksgroep dan in het hoger onderwijs. Dit suggereert volgens hen dat de feedbackvorm EF wellicht minder effectief is voor leerlingen in het basisonderwijs. Dit zou een verklaring kunnen zijn voor het feit dat de resultaten binnen dit onderzoek niet significant van elkaar verschillen. Verder geven van der Kleij et al. (2015) aan dat EF vooral een groter effect heeft wanneer het om opdrachten op een complexer denkniveau gaat. De opdrachten en vragen die tijdens dit onderzoek centraal stonden, richten zich echter op een lager denkniveau. Ook dit zou invloed gehad kunnen hebben op het uitblijven van significante verschillen binnen dit onderzoek.

Sterke punten en beperkingen

Dit onderzoek kent zowel sterke punten als beperkingen. Het eerste sterke punt is dat de resultaten van de KCR en EF groep werden afgezet tegen die van een controlegroep (KR). Daarbij werden de leerlingen willekeurig over de drie onderzoeksgroepen verdeeld. Door deze willekeurige verdeling zaten leerlingen uit verschillende klassen bij elkaar in een onderzoeksgroep en speelden leerlingen uit verschillende onderzoeksgroepen tegelijk de serious game. Hiermee werd voorkomen dat de omstandigheden tijdens het spelen van de game van invloed zouden zijn op de resultaten. Bovendien speelden alle leerlingen Plantendetective op maandagochtend en werd de nameting voor alle groepen 's middags afgenomen. Tot slot vulden de leerlingen niet alleen een kennistest in voor de nameting, maar vond na twee weken ook een retentiemeting plaats. Girard et al. (2013) geven namelijk aan dat het belangrijk is om te meten of de verworven kennis tijdens een serious game ook op de lange termijn behouden blijft.

Naast deze sterke punten kent dit onderzoek ook een aantal beperkingen. Een van de tekortkomingen van dit onderzoek is dat de participanten allemaal op dezelfde basisschool zitten. Hier was voor gekozen omdat de invloed van externe factoren op deze manier zo beperkt mogelijk blijft. Een nadeel is echter dat dit de generaliseerbaarheid van de bevindingen in het onderzoek beperkt, zeker gezien het bovengemiddeld hoge uitstroomniveau van de leerlingen op de participerende school.

Een andere limitatie van dit onderzoek is de manier waarop de kennis van de leerlingen op het gebied van natuuronderwijs is gemeten. Hiervoor zijn drie kennistesten ontwikkeld: een voor-, na- en retentiemeting. Om de kennistesten kort te houden en op deze manier zo min mogelijk tijd van de leerlingen en leerkrachten te vragen, bevatte elke kennistest slechts tien vragen. Volgens van der Kleij et al. (2015) is het echter belangrijk dat de voor- en nameting voldoende items hebben om gegronde uitspraken te kunnen doen over de leereffecten. Tien vragen per kennistest is aan de lage kant,

waardoor het lastig is om duidelijke verschillen tussen de metingen te kunnen waarnemen. Uit de betrouwbaarheidsmeting blijkt bovendien dat de drie kennistesten een lage tot matige betrouwbaarheid hebben. Ook voor de kwaliteitsmeting is een betrouwbaarheidsanalyse uitgevoerd. Vier van de vijf onderdelen uit de kwaliteitsmeting bleken voldoende betrouwbaar te zijn. Alleen de betrouwbaarheid van de vragen uit de NASA-TLX (Hart & Staveland, 1988) bleek matig te zijn. Dit resultaat is opvallend omdat de NASA-TLX doorgaans wel een goede betrouwbaarheid heeft (Longo, 2018). Mogelijk is de betrouwbaarheid in dit onderzoek lager omdat de vragen niet zijn geoptimaliseerd voor basisschoolleerlingen.

Op basis van het onderzoek van Mason en Bruning (2001) werd besloten om voor dit onderzoek gebruik te maken van directe *EF response contingent* feedback, omdat dit de beste feedbackvorm zou zijn voor leerlingen met een laag prestatieniveau. Hierbij was de verwachting dat de participerende leerlingen binnen deze doelgroep zouden vallen, aangezien zij nog op de basisschool zitten. Uit de demografische gegevens blijkt echter dat de meeste leerlingen een bovengemiddeld hoog niveau hebben. Leerlingen met een hoog niveau zouden evenals leerlingen met een laag niveau het best presteren met *EF response contingent* feedback, maar in tegenstelling tot leerlingen met een laag niveau zouden zij meer baat hebben bij uitgestelde feedback (Shute, 2008). Daarnaast geven Mason en Bruning (2001) aan dat leerlingen met een hoog niveau meer zouden hebben aan minder specifieke feedback waarbij zij de ruimte krijgen om zelf actief over de leerstof na te denken. Dit betekent dat bij dit onderzoek mogelijk niet de meest ideale vorm van EF is toegepast.

Tot slot geven van der Kleij et al. (2015) in hun meta-analyse aan dat bij een groot aantal onderzoeken aangenomen wordt dat de leerlingen aandacht besteden aan de aangeboden feedback, terwijl recent onderzoek (van der Kleij et al., 2012) uitwijst dat sommige leerlingen geschreven feedback in een computergestuurde omgeving gemakkelijk negeren. Binnen dit onderzoek is niet expliciet gemonitord of de leerlingen de feedback in Plantendetective daadwerkelijk hebben gelezen. Van der Kleij et al. (2011) geven aan dat sommige onderzoeken de sessieduur van de game meten om te onderzoeken of de participanten de feedback lezen. De gemiddelde sessieduur van de serious game Plantendetective lijkt iets langer te zijn voor de EF groep dan voor de KR en KCR groep. Dit suggereert dat de leerlingen in de EF groep langer de tijd hebben genomen om Plantendetective te spelen, wat het aannemelijk maakt dat zij de feedback ook daadwerkelijk hebben gelezen (van der Kleij et al., 2011). De verschillen in sessieduur tussen de drie onderzoeksgroepen zijn echter niet significant en zouden ook kunnen worden toegeschreven aan andere factoren, zoals het studietempo van de leerling. Bovendien is binnen elke experimentele groep veel variatie in de sessieduur. Hierdoor is het onmogelijk te zeggen of de leerlingen daadwerkelijk gebruik hebben gemaakt van de feedback.

Aanbevelingen voor vervolgonderzoek

Dit onderzoek draagt bij aan de kennis over de inzet van feedback binnen serious games. Deze kennis is van belang aangezien serious games steeds populairder worden in het (basis)onderwijs, maar deze games niet altijd even effectief zijn voor het verbeteren van de leerresultaten van leerlingen. Effectief gebruik van feedback binnen serious games kan het leerrendement van de game verhogen. Dit maakt de inzet van serious games in de onderwijspraktijk aantrekkelijker, omdat leerlingen op deze manier zelfstandiger met serious games kunnen leren. Daarom is het belangrijk dat voldoende kennis verworven wordt over de effectiviteit van verschillende feedbackvormen binnen serious gaming in het basisonderwijs.

Aangezien op basis van dit onderzoek nog geen duidelijke conclusies kunnen worden getrokken, is het aan te raden om verder onderzoek te doen naar de effectiviteit van verschillende feedbackvormen in serious games binnen het basisonderwijs. Bij dit onderzoek is slechts met één basisschool gewerkt. Om de generaliseerbaarheid van de bevindingen te verhogen, is het van belang om bij vervolgonderzoek participanten van verschillende basisscholen te selecteren. Hierbij dient wel rekening te worden gehouden met externe factoren die van invloed kunnen zijn op de resultaten, zoals de geografische ligging van de school en het uitstroomprofiel. Door met meerdere basisscholen te werken kan het aantal participanten worden verhoogd. Dit maakt het onder andere mogelijk om statistische analyses over het prestatieniveau van de leerlingen uit te voeren. Hierdoor kan worden onderzocht of het prestatieniveau van de leerling invloed heeft op de effectiviteit van de feedbackvormen.

Om de leereffecten op een juiste manier te kunnen meten, is het belangrijk dat goede meetinstrumenten worden ontwikkeld. Het is daarom aan te raden om bij vervolgonderzoek gebruik te maken van een gevalideerde kennistest om de leereffecten te kunnen meten. Dit betekent dat de serious game over een onderwerp zal moeten gaan waarvan al gevalideerde meetinstrumenten beschikbaar zijn. Wanneer dit niet het geval is, zal een eigen meetinstrument ontwikkeld moeten worden dat over voldoende vragen beschikt om duidelijke verschillen tussen de metingen te kunnen constateren. Daarnaast dient de betrouwbaarheid van het onderdeel taakbelasting binnen de kwaliteitsmeting te worden verbeterd.

Aangezien niet met zekerheid te zeggen is of de participanten de feedback binnen de serious game daadwerkelijk gelezen hebben, zou dit bij vervolgonderzoek nader onderzocht kunnen worden. Volgens van der Kleij et al. (2015) zou dit bijvoorbeeld kunnen door de tijd te meten dat de feedback in beeld weergegeven wordt. Met behulp van eye-tracking technologieën zou nog gedetailleerder vastgelegd kunnen worden hoe de leerlingen omgaan met de geboden feedback binnen computergestuurde omgevingen. Bij een groot aantal participanten kan het echter lastig zijn om iedereen van eye-tracking apparatuur te voorzien. In dat geval is het ook mogelijk om kwalitatieve

data te verzamelen door de leerlingen na afloop van het onderzoek te vragen of ze de feedback gelezen hebben en hier gebruik van hebben gemaakt.

Conclusie

Op basis van dit onderzoek kan worden geconcludeerd dat er geen verschil is in het effect van de feedbackvormen KR, KCR en EF op het leereffect van een serious game bij leerlingen op de basisschool. Dit resultaat staat in contrast met eerder onderzoek in het middelbaar- en hoger onderwijs waaruit blijkt dat EF de meest effectieve feedbackvorm is bij serious gaming (Jaehnig & Miller, 2007; Mason & Bruning, 2001; van der Kleij et al., 2015). Vervolgonderzoek bij meerdere scholen, met meer geavanceerde meetinstrumenten voor het meten en interpreteren van de leerresultaten en de toepassing van technieken om te onderzoeken of de leerlingen de feedback daadwerkelijk lezen, zal uit moeten wijzen of EF ook voor leerlingen in het basisonderwijs de meest effectieve feedbackvorm is.

Referenties

- Abdul Jabbar, A. I., & Felicia, P. (2015). Gameplay Engagement and Learning in Game-Based Learning: A Systematic Review. *Review of Educational Research*, 85(4), 740-779. doi:10.3102/0034654315577210
- Annetta, L. A. (2010). The "I's" Have It: A Framework for Serious Educational Game Design. *Review of General Psychology*, 14(2), 105-112. doi:10.1037/a0018985
- Annetta, L. A., Minogue, J., Holmes, S. Y., & Cheng, M.-T. (2009). Investigating the impact of video games on high school students' engagement and learning about genetics. *Computers & Education*, 53(1), 74-85. doi:10.1016/j.compedu.2008.12.020
- Barab, S., & Dede, C. (2007). Games and Immersive Participatory Simulations for Science Education: An Emerging Type of Curricula. *Journal of Science Education and Technology*, 16(1), 1-3. doi:10.1007/s10956-007-9043-9
- Bellotti, F., Kapralos, B., Lee, K., Moreno-Ger, P., & Berta, R. (2013). Assessment in and of Serious Games: An Overview. *Advances in Human-Computer Interaction*, 2013, 1-11. doi:10.1155/2013/136864
- Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital Games, Design, and Learning: A Systematic Review and Meta-Analysis. *Review of Educational Research*, 86(1), 79-122. doi:10.3102/0034654315582065
- Clark, R. E., Yates, K., Early, S., & Moulton, K. (2010). An Analysis of the Failure of Electronic Media and Discovery-Based Learning. *Handbook of Improving Performance in the Workplace: Instructional Design and Training Delivery*. doi:10.1002/9780470587089.ch8
- Cohen, J. (1988). *Statistical power ANALYSIS for the Behavioral sciences* (Vol. 2nd). Hillsdale, NJ: Erlbaum.
- Divjak, B., & Tomić, D. (2011). *The impact of game-based learning on the achievement of learning goals and motivation for learning mathematics - Literature review* (Vol. 35).
- Dunbar, N. E., Jensen, M. L., Miller, C. H., Bessarabova, E., Lee, Y.-H., Wilson, S. N., . . . Schuetzler, R. M. (2017). Mitigation of Cognitive Bias with a Serious Game: Two Experiments Testing Feedback Timing and Source. *International Journal of Game-Based Learning (IJGBL)*, 7(4), 86-100. doi:10.4018/IJGBL.2017100105
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving Students' Learning With Effective Learning Techniques: Promising Directions From Cognitive and Educational Psychology. *Psychological Science in the Public Interest*, 14(1), 4-58. doi:10.1177/1529100612453266
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175-191. doi:10.3758/BF03193146
- Giannakas, F., Kambourakis, G., Papasalouros, A., & Gritzalis, S. (2018). A critical review of 13 years of mobile game-based learning. *Educational Technology Research and Development*, 66(2), 341-384. doi:10.1007/s11423-017-9552-z
- Girard, C., Ecalle, J., & Magnan, A. (2013). Serious games as new educational tools: how effective are they? A meta-analysis of recent studies. *Journal of Computer Assisted Learning*, 29(3), 207-219. doi:10.1111/j.1365-2729.2012.00489.x
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock & N. Meshkati (Eds.), *Advances in Psychology* (Vol. 52, pp. 139-183): North-Holland.
- Harteveld, C., Guimarães, R., Mayer, I. S., & Bidarra, R. (2010). Balancing Play, Meaning and Reality: The Design Philosophy of LEVEE PATROLLER. *Simulation & Gaming*, 41(3), 316-340. doi:10.1177/1046878108331237
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Londen, Engeland: Routledge.

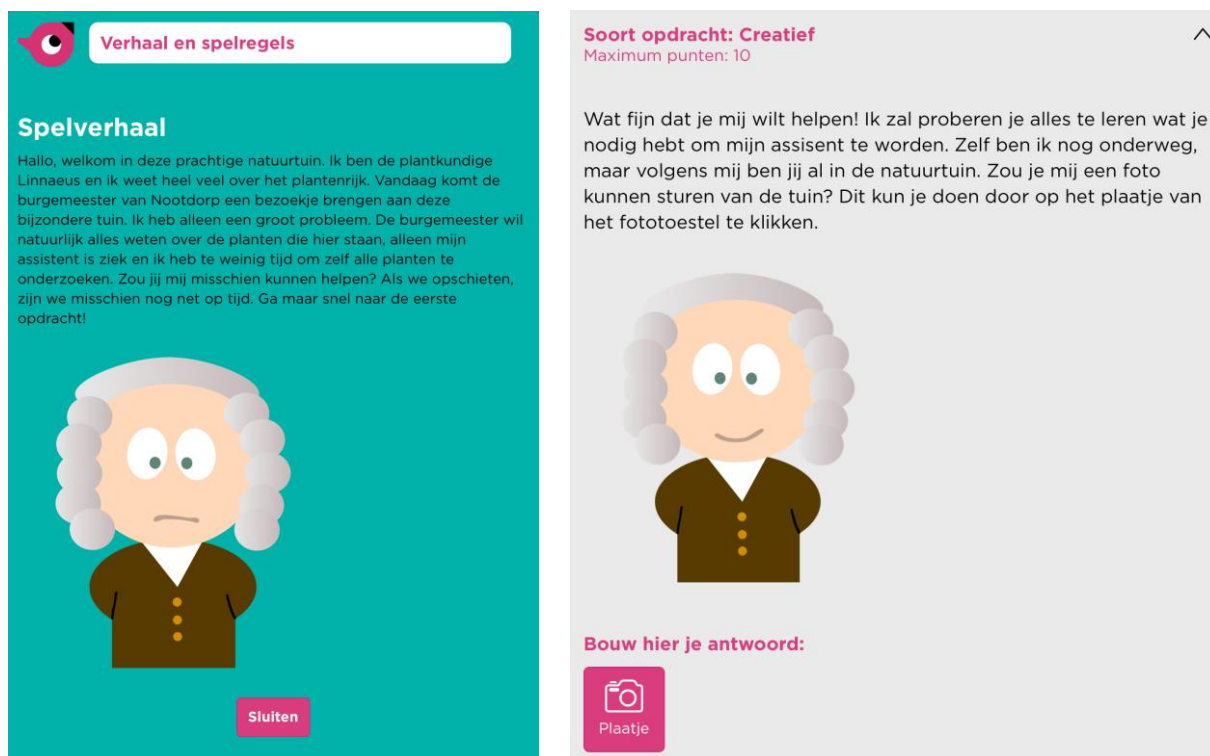
- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81-112. doi:10.3102/003465430298487
- Huizenga, J., Admiraal, W., Akkerman, S., & ten Dam, G. (2009). Mobile game-based learning in secondary education: engagement, motivation and learning in a mobile city game. *Journal of Computer Assisted Learning*, 25(4), 332-344. doi:10.1111/j.1365-2729.2009.00316.x
- Jaehnig, W., & Miller, M. L. (2007). Feedback types in programmed instruction: A systematic review. *PSYCHOLOGICAL RECORD*, 57(2), 219-232. doi:10.1007/BF03395573
- Kiili, K. (2005). Digital game-based learning: Towards an experiential gaming model. *The Internet and Higher Education*, 8(1), 13-24. doi:10.1016/j.iheduc.2004.12.001
- Lai, N. K., Ang, T. F., Por, L. Y., & Liew, C. S. (2018). The impact of play on child development - a literature review. *European Early Childhood Education Research Journal*, 26(5), 625-643. doi:10.1080/1350293X.2018.1522479
- Li, M.-C., & Tsai, C.-C. (2013). Game-Based Learning in Science Education: A Review of Relevant Research. *Journal of Science Education and Technology*, 22(6), 877-898. doi:10.1007/s10956-013-9436-x
- Longo, L. (2018). *On the Reliability, Validity and Sensitivity of Three Mental Workload Assessment Techniques for the Evaluation of Instructional Designs: A Case Study in a Third-level Course*. Paper presented at the CSEDU 2018.
- Mason, B. J., & Bruning, R. (2001). *Providing Feedback in Computer-based Instruction: What the Research Tells Us*. Center for Instructional Innovation, University of Nebraska-Lincoln.
- McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger, H. L. (2011). Test-Enhanced Learning in a Middle School Science Classroom: The Effects of Quiz Frequency and Placement. *Journal of Educational Psychology*, 103(2), 399-414. doi:10.1037/a0021782
- Nadolski, R. J., & Hummel, H. G. K. (2017). Retrospective cognitive feedback for progress monitoring in serious games. *British Journal of Educational Technology*, 48(6), 1368-1379. doi:10.1111/bjet.12503
- Narciss, S., & Huth, K. (2004). How to design informative tutoring feedback for multi-media learning. In H. M. Niegemann, D. Leutner, & R. Brünken (Eds.), *Instructional Design for Multimedia learning* (pp. 181-195). Münster, New York: Waxman.
- Papanastasiou, G., Drigas, A., & Skianis, C. (2017). Serious Games in Preschool and Primary Education: Benefits And Impacts on Curriculum Course Syllabus. *International Journal of Emerging Technologies in Learning (iJET)*, 12(1), 44. doi:10.3991/ijet.v12i01.6065
- Perrotta, C., Featherstone, G., Aston, H., & Houghton, E. (2013). *Game-based learning: latest evidence and future directions*. Slough: National Foundation for Educational Research.
- Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of Game-Based Learning. *Educational Psychologist*, 50(4), 258-283. doi:10.1080/00461520.2015.1122533
- Rheinberg, F., Vollmeyer, R., & Engeser, S. (2003). The assessment of flow. In J. Stiensmeier-Pelster & F. Rheinberg (Eds.), *Diagnosis of motivation and self-concept* (pp. 261-279). Göttingen: Hogrefe.
- Roediger, H. L., & Pyc, M. A. (2012). Inexpensive techniques to improve education: Applying cognitive psychology to enhance educational practice. *Journal of Applied Research in Memory and Cognition*, 1(4), 242-248. doi:10.1016/j.jarmac.2012.09.002
- Rogers, R. (2017). The motivational pull of video game feedback, rules, and social interaction: Another self-determination theory approach. *Computers in Human Behavior*, 73, 446-450. doi:10.1016/j.chb.2017.03.048
- Rohrer, D., & Pashler, H. (2010). Recent Research on Human Learning Challenges Conventional Instructional Strategies. *Educational Researcher*, 39(5), 406-412. doi:10.3102/0013189X10374770
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68-78. doi:10.1037/0003-066X.55.1.68
- Shute, V. J. (2008). Focus on Formative Feedback. *Review of Educational Research*, 78(1), 153-189. doi:10.3102/0034654307313795

- Tsai, C.-C., Tsai, F.-H., & Lin, K.-Y. (2015). The evaluation of different gaming modes and feedback types on game-based formative assessment in an online learning environment. *Computers & Education*, 81, 259-269. doi:10.1016/j.compedu.2014.10.013
- van der Kleij, F. M., Eggen, T. J. H. M., Timmers, C. F., & Veldkamp, B. P. (2012). Effects of feedback in a computer-based assessment for learning. *Computers & Education*, 58(1), 263-272. doi:10.1016/j.compedu.2011.07.020
- van der Kleij, F. M., Feskens, R. C. W., & Eggen, T. J. H. M. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: a meta-analysis. *Review of Educational Research*, 85(4), 475-511. doi:10.3102/0034654314564881
- van der Kleij, F. M., Timmers, C., & Eggen, T. (2011). *The effectiveness of methods for providing written feedback through a computer-based assessment for learning: A systematic review* (Vol. 19).
- Wilson, C. D., Reichsman, F., Mutch-Jones, K., Gardner, A., Marchi, L., Kowalski, S., . . . Dorsey, C. (2018). Teacher Implementation and the Impact of Game-Based Science Curriculum Materials. *Journal of Science Education and Technology*, 27(4), 285-305. doi:10.1007/s10956-017-9724-y
- Wouters, P., Nimwegen, v. C., Oostendorp, v. H., & Spek, v. d. E. D. E. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, 105(2), 249-265. doi:10.1037/a0031311
- Wrzesien, M., & Alcañiz Raya, M. (2010). Learning in serious virtual worlds: Evaluation of learning effectiveness and appeal to students in the E-Junior project. *Computers & Education*, 55(1), 178-187. doi:10.1016/j.compedu.2010.01.003
- Young, M. F., Slota, S., Cutter, A. B., Jalette, G., Mullin, G., Lai, B., . . . Yukhymenko, M. (2012). Our Princess Is in Another Castle: A Review of Trends in Serious Gaming for Education. *Review of Educational Research*, 82(1), 61-89. doi:10.3102/0034654312436980

Bijlage 1. Verantwoording serious game volgens het framework van Annetta (2010)

Bij het ontwerpen van Plantendetective, de serious game die binnen dit onderzoek centraal staat, vormde het framework van Annetta (2010) de basis. In deze bijlage wordt eerst een algemeen beeld gegeven van de serious game waarna voor elk van de zes elementen uit het framework wordt uitgewerkt hoe dit element in de serious game vertaald is.

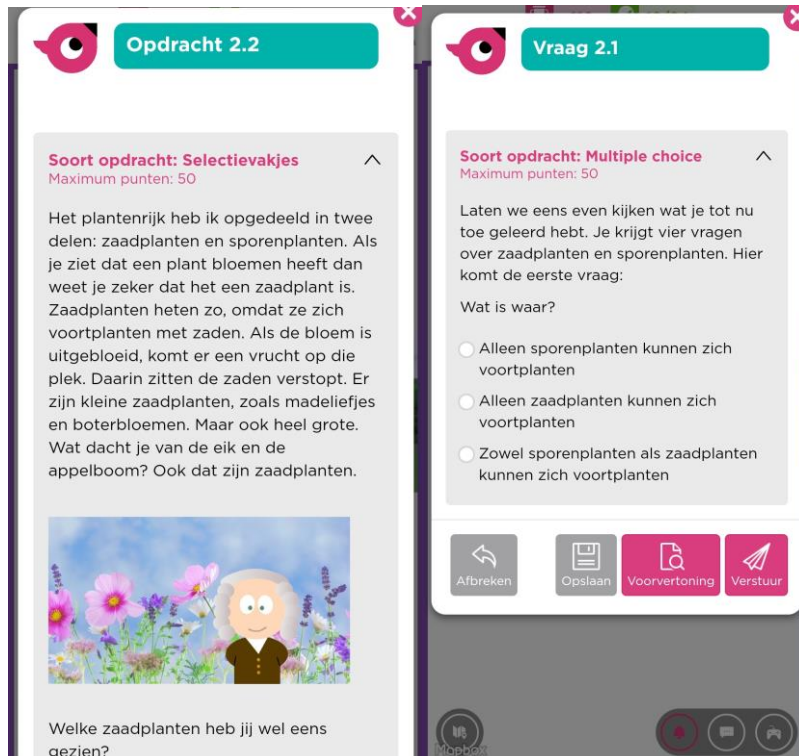
De serious game Plantendetective is ontworpen voor het vak natuuronderwijs en gaat over het plantenrijk. Doel van de game is dat de leerlingen leren hoe zij het plantenrijk kunnen indelen, bijvoorbeeld in zaadplanten en sporenplanten. Hierbij krijgen de leerlingen hulp van het spelpersonage Linnaeus, de plantkundige die de indeling van het plantenrijk heeft bedacht. Plantendetective begint met een speluitleg waarbij Linnaeus vertelt dat de burgemeester een bezoek komt brengen aan de natuurtuin van de school. Voordat de burgemeester komt, moet Linnaeus alle planten nog onderzoeken. Zijn assistent is alleen ziek, dus vraagt hij de leerling om hulp (zie Figuur 1).



Figuur 1. Speluitleg en eerste opdracht in de serious game

De leerlingen bevinden zich op het moment van spelen in de natuurtuin van de school. Doordat het een mobiele serious game betreft, kunnen de leerlingen de game op hun tablet spelen en zijn ze hierbij niet gebonden aan het klaslokaal. De leerlingen kunnen opdrachten vervullen met behulp van tekst, de camera- of audiofunctie van de tablet. Linnaeus geeft de leerlingen steeds informatie voordat zij een

opdracht krijgen. Vervolgens wordt de kennis van de leerling getest met een meerkeuzevraag. Een voorbeeld van deze informatie en kennisvragen is te zien in Figuur 2.



Figuur 2. Voorbeeld van informatie die Linnaeus aan de leerling geeft (links) en een meerkeuzevraag om de opgedane kennis te toetsen (rechts)

De game Plantendetective bestaat uit drie levels (zie Figuur 3). In het eerste level vertelt Linnaeus over het verschil tussen het plantenrijk en het schimmelrijk. Hierbij geeft hij eerst informatie en daarna krijgen de leerlingen een opdracht. Linnaeus geeft steeds hints en helpt de leerlingen zoveel mogelijk. Als de leerling de opdracht niet begrijpt of extra aanwijzingen nodig heeft, kan hij via de chatbox contact opnemen met Linnaeus. Het eerste level wordt afgesloten met een viertal meerkeuzevragen die gaan over alle kennis die in het eerste level aan bod is gekomen. Voor elk goede antwoord kan de leerling 50 punten verdienen. De leerling krijgt, afhankelijk van de onderzoeksgroep waar hij in zit, KR, KCR of EF *response contingent* feedback op zijn antwoorden op deze vragen (zie Figuur 4). De feedback wordt direct na het beantwoorden van de vraag gegeven. De leerling verdient zodra hij alle vier de vragen heeft beantwoord, ongeacht het aantal punten dat hij hierbij heeft gehaald, een badge waarmee hij officieel assistent wordt.

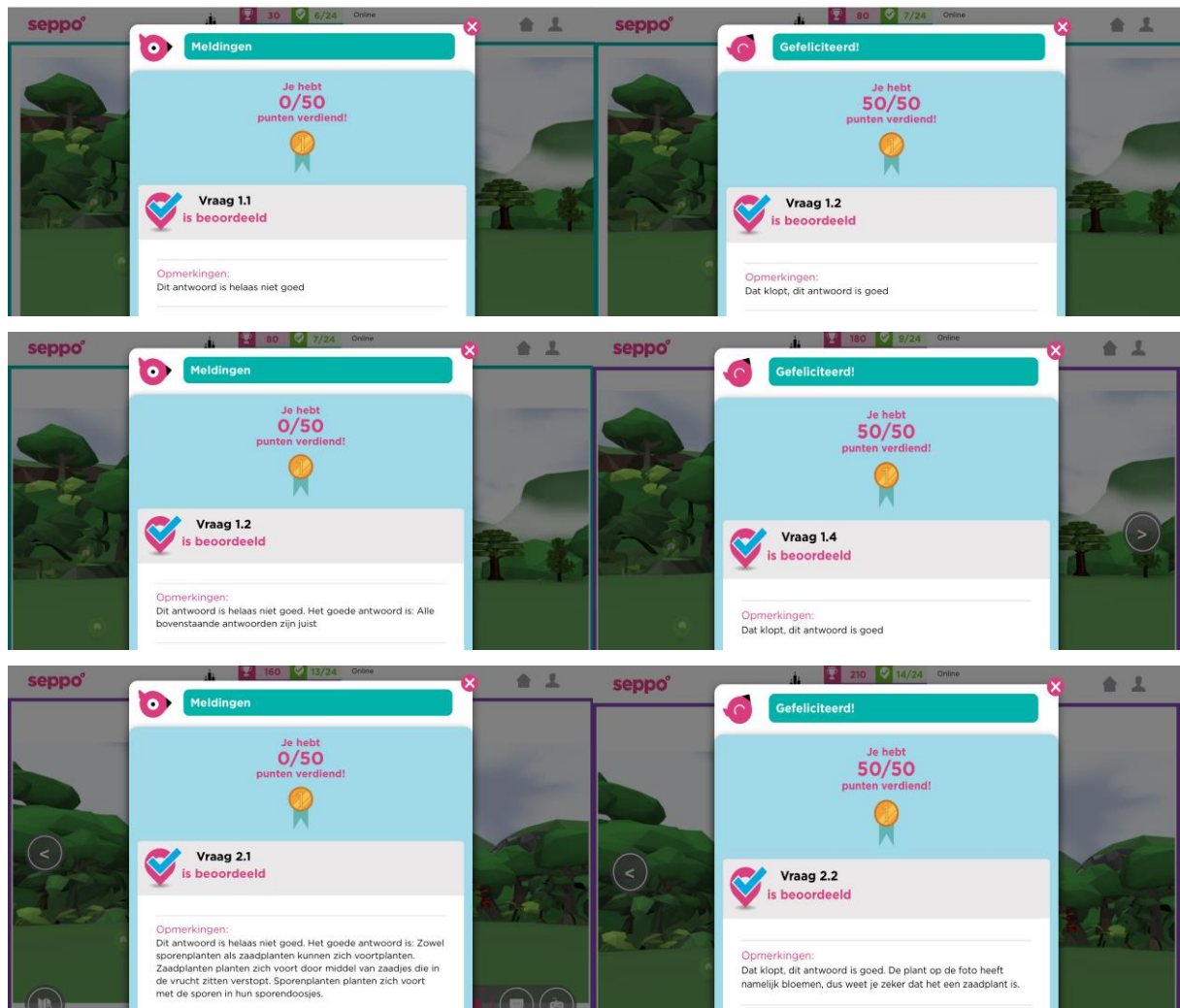
In het tweede level legt Linnaeus uit dat het plantenrijk kan worden ingedeeld in zaadplanten en sporenplanten. De leerling krijgt opnieuw een aantal opdrachten in de natuur tuin. Omdat hij nu officieel de assistent van Linnaeus is, krijgt de leerling hierbij minder begeleiding aangeboden dan in

het eerste level. Het tweede level wordt opnieuw afgesloten met vier meerkeuzevragen waarna de leerling de badge ‘hoofdassistent’ verdient.

Tot slot gaat het derde level van de serious game in op het verschil tussen bedektzadige- en naaktzadige planten. Het level start met een uitleg van Linnaeus, maar daarna krijgt de leerling geen extra hulp meer bij het uitvoeren van de opdrachten. Dit laatste level eindigt wederom met een meerkeuzetest van vier vragen over de lesstof. Als de leerling deze vier vragen heeft doorlopen, krijgt hij de badge ‘plantendetective’ en is het spel afgelopen.



Figuur 3. Startscherm en verdeling van de verschillende levels



Figuur 4. Verschillen in feedback voor KR, KCR en EF (van boven naar beneden)

Identity

Een goed spel is in staat de speler te laten geloven dat hij een uniek individu is binnen de omgeving. In videogames wordt de identiteit van de speler vaak gerepresenteerd door een avatar (Annetta, 2010). In de game Plantendetective is het echter niet mogelijk om gebruik te maken van avatars. Om de leerling toch het gevoel te geven dat hij een uniek individu is binnen de serious game, start de game met een inleiding waarbij het spelpersonage, de plantkundige Linnaeus, zichzelf voorstelt en de hulp van de leerling vraagt. Op deze manier krijgt de leerling een belangrijke taak toegewezen als assistent van Linnaeus en is zijn identiteit bij de start van de game meteen duidelijk, waardoor hij het gevoel krijgt onderdeel te zijn van het spel.

Immersion

Immersion in een game wordt bereikt doordat de speler betrokken is bij de content van het spel en intrinsiek gemotiveerd is om het doel te behalen. Hierbij spelen uitdaging, nieuwsgierigheid, controle en het doel van het spel een rol. Wanneer de leerling betrokken en gemotiveerd genoeg is, kan hij in een *state of flow* terechtkomen (Annetta, 2010). Om de kans dat de leerling in deze *state of flow* komt te bevorderen, wordt Plantendetective gespeeld in de natuurtuin van de school. Deze omgeving is onderdeel van de serious game waardoor de leerling zich optimaal op het spel kan concentreren. Door de serious game te starten met de speluitleg, wordt het doel van de game meteen duidelijk. Verder heeft de leerling veel controle over de opdrachten. Hij mag zelf kiezen op welke manier hij de opdrachten uitvoert. Tot slot geeft Kiili (2005) aan dat geschikte feedback van groot belang is om in een *state of flow* te kunnen komen. Aangezien de feedback zo snel mogelijk na het uitvoeren van de opdracht moet worden gegeven (Annetta, 2010), krijgen de leerlingen hun feedback direct na het indienen van het antwoord te zien. Om de motivatie te verhogen, kunnen de leerlingen naast deze feedback ook punten verdienen met hun antwoorden op de vragen.

Interactivity

Communicatie is belangrijk binnen een game. Het maakt hierbij niet uit of het gaat om communicatie met andere spelers of met een computergestuurd personage uit het spel (Annetta, 2010). Aangezien contact met medespelers invloed zou kunnen hebben op de onderzoeksresultaten, kunnen de leerlingen binnen de game Plantendetective alleen met het spelpersonage Linnaeus communiceren. Dit kunnen zij doen door gebruik te maken van de chatbox. In deze chatbox kunnen de leerlingen verduidelijkende vragen aan Linnaeus stellen of opmerkingen plaatsen. Deze vragen worden door de onderzoeker beantwoord aan de hand van een vast format.

Increasing complexity

Goede games bestaan vaak uit meerdere levels die kunnen zorgen voor een oplopende complexiteit van de concepten en content in de serious game. Het doel van de ontwerper is hierbij om steeds weer voor een balans in plezier en uitdaging te zorgen (Annetta, 2010). Plantendetective is daarom opgebouwd uit drie levels in oplopende moeilijkheid. In het eerste level krijgt de leerling nog veel uitleg en aanwijzingen van Linnaeus. In het tweede level bouwt deze ondersteuning langzaam af en het derde level moet de leerling vrijwel zonder hulp kunnen uitvoeren. Aan het einde van ieder level is een korte test opgenomen om de opgedane kennis binnen het level te meten. Pas als de leerling alle vragen heeft beantwoord, verdient hij een badge en mag hij door naar het volgende level.

Informed teaching

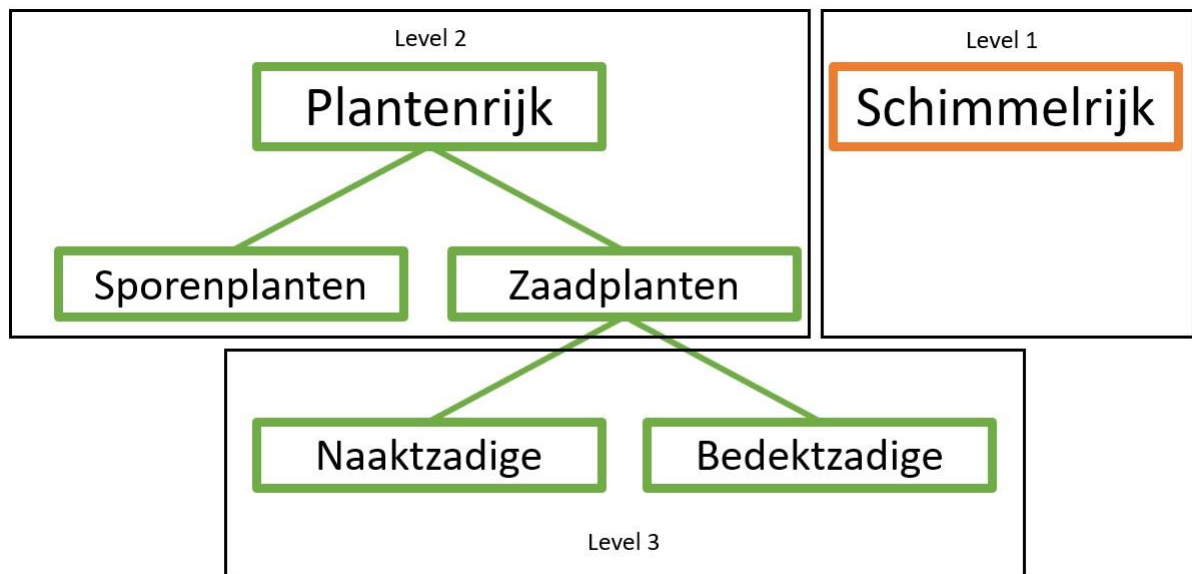
Informed teaching heeft betrekking op de manier waarop feedback en toetsing zijn ingebouwd in de serious game. Door de data die tijdens het spelen van het spel opgeslagen wordt, is het mogelijk om de kennis van leerlingen te beoordelen (Annetta, 2010). Gedurende het experiment worden de handelingen van de leerling binnen de game Plantendetective gevolgd. Hierbij worden de antwoorden die hij geeft op de meerkeuzevragen opgeslagen en voorzien van passende, computergestuurde feedback. Ook andere handelingen binnen het spel, zoals het versturen van foto's en de communicatie met spelpersonage Linnaeus, worden opgeslagen en kunnen achteraf beoordeeld worden.

Instructional

Om een serious game succesvol toe te kunnen passen in de onderwijspraktijk, is het belangrijk dat hij goed aansluit op de andere lessen (Annetta, 2010). De game Plantendetective is daarom opgesteld aan de hand van de leerdoelen en lesinhoud van Argus Clou, de lesmethode die op de school van de deelnemende leerlingen voor de lessen natuuronderwijs wordt gebruikt. In de serious game staan de volgende leerdoelen centraal:

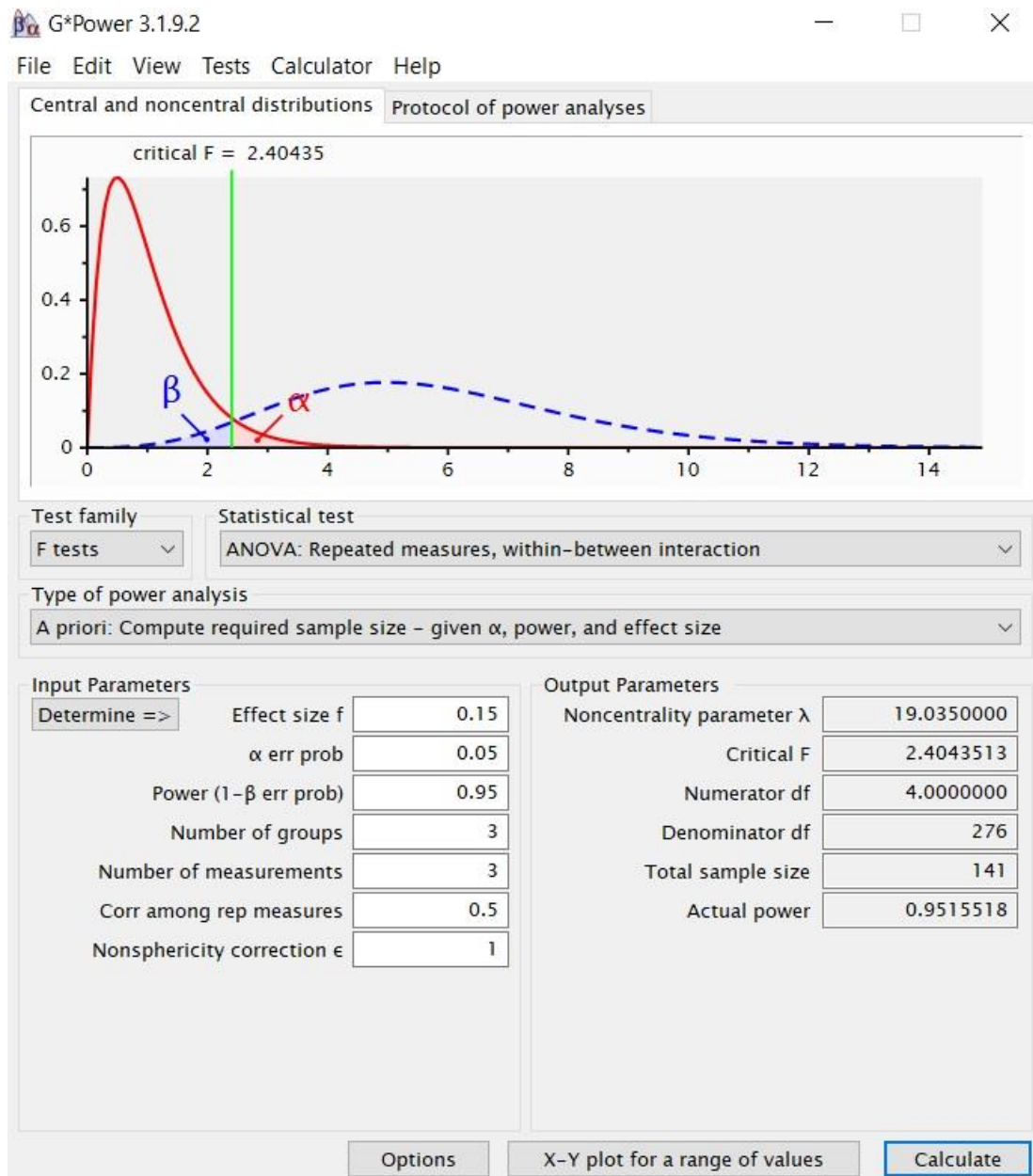
- Leerlingen kunnen zaadplanten en sporenplanten van elkaar onderscheiden
- Leerlingen kunnen uitleggen wat de verschillen tussen schimmels en planten zijn
- Leerlingen kunnen de kenmerken van bedektzadige en naaktzadige planten noemen

Bij het begin van ieder level krijgt de leerling nieuwe informatie aangeboden. Hierbij wordt eerst de voorkennis geactiveerd. Aangezien visualisatie een krachtige cognitieve strategie is (Annetta, 2010), wordt tekstuele informatie aangevuld met afbeeldingen en filmpjes. De verdeling van de leerstof over de verschillende levels is terug te vinden in Figuur 5.



Figuur 5. Aangeboden onderwerpen in de verschillende levels van het spel Plantendetective

Bijlage 2. Berekening G*Power



Bijlage 3. Kwaliteitsmeting game

Het ontwerpen van de serious game Plantendetective is gebeurd aan de hand van het framework van Annetta (2010). Dit framework bestaat uit zes elementen die samen voor een goede game zorgen. Het eerste element, *identity*, gaat over het gevoel van de leerling dat hij een uniek individu is binnen de game. Daarna volgt *immersion*, de mate van onderdompeling in de game. Hierbij zijn de intrinsieke motivatie van de leerling en de *state of flow* van belang. Om in een *state of flow* te komen, moet een leerling zich betrokken voelen bij de game. Een goede game heeft ook een bepaalde mate van *interactivity*, waarbij de leerling kan communiceren met medespelers of een computergestuurd personage. Vervolgens is het bij de *increasing complexity* van de game belangrijk dat de leerling steeds de juiste mate van plezier en uitdaging ervaart. *Informed teaching* gaat over de manier waarop feedback en toetsing in de game zijn geïntegreerd. Tot slot is de lesinhoud (*instructional*) belangrijk voor een goede game (Annetta, 2010). Om de kwaliteit van de game Plantendetective te kunnen meten, werden de zes elementen achteraf door de leerlingen geëvalueerd door middel van een kwaliteitsmeting. De vragen uit deze vragenlijst staan hieronder per element geordend.

Identity

Binnen de gevalideerde vragenlijsten waren geen vragen opgenomen om het onderdeel *identity* te kunnen meten. Daarom zijn onderstaande vragen door de onderzoeker zelf ontwikkeld.

1. Ik had het gevoel onderdeel te zijn van de game
[1: Geheel mee oneens – 5: Geheel mee eens]
2. Ik vind dat ik een belangrijke taak vervulde binnen de game
[1: Geheel mee oneens – 5: Geheel mee eens]

Immersion

Voor het meten van *immersion* werden verschillende gevalideerde vragenlijsten gebruikt. Met behulp van de Intrinsic Motivation Inventory (IMI) (Ryan & Deci, 2000) werd de motivatie van leerlingen gemeten. Hierbij is gekozen voor het gebruik van de sub schalen *interest / enjoyment* en *effort / importance*. De sub schaal *interest / enjoyment* vormt de basis van de IMI en wordt gezien als het belangrijkste meetinstrument voor de intrinsieke motivatie. De *effort / importance* sub schaal laat daarnaast zien hoe hard de leerling zijn best heeft gedaan in de game.

1. Ik heb erg genoten van het doen van deze game
[1: Geheel mee oneens – 7: Geheel mee eens]

2. Deze game was leuk om te doen
[1: Geheel mee oneens – 7: Geheel mee eens]
3. Ik denk dat het doen van deze game saai was*
[1: Geheel mee oneens – 7: Geheel mee eens]
4. Deze game heeft mijn aandacht helemaal niet getrokken*
[1: Geheel mee oneens – 7: Geheel mee eens]
5. Ik zou deze game als zeer interessant willen omschrijven
[1: Geheel mee oneens – 7: Geheel mee eens]
6. Toen ik deze game deed, dacht ik hoe plezierig het was
[1: Geheel mee oneens – 7: Geheel mee eens]
7. Ik heb veel inspanning geleverd voor deze game
[1: Geheel mee oneens – 7: Geheel mee eens]
8. Ik heb niet hard mijn best gedaan om deze game goed te doen*
[1: Geheel mee oneens – 7: Geheel mee eens]
9. Ik heb heel hard mijn best gedaan in deze game
[1: Geheel mee oneens – 7: Geheel mee eens]
10. Het was belangrijk voor me om de game goed te doen
[1: Geheel mee oneens – 7: Geheel mee eens]
11. Ik heb weinig energie in de game gestoken*
[1: Geheel mee oneens – 7: Geheel mee eens]

De *state of flow* werd vervolgens gemeten met de Flow-short (Rheinberg et al., 2003).

1. Ik voelde me prettig uitgedaagd
[1: Geheel niet waar – 7: Helemaal waar]
2. Mijn gedachten verliepen soepel
[1: Geheel niet waar – 7: Helemaal waar]
3. Mijn hoofd was helemaal helder
[1: Geheel niet waar – 7: Helemaal waar]
4. Ik was helemaal verdiept in wat ik deed
[1: Geheel niet waar – 7: Helemaal waar]
5. De juiste gedachten kwamen vanzelf
[1: Geheel niet waar – 7: Helemaal waar]
6. Bij iedere stap wist ik wat ik moest doen
[1: Geheel niet waar – 7: Helemaal waar]

7. Ik had het gevoel dat ik alles onder controle had
[1: Geheel niet waar – 7: Helemaal waar]
8. Ik was in gedachten verzonken
[1: Geheel niet waar – 7: Helemaal waar]
9. Ik had er geen idee van hoe snel de tijd liep
[1: Geheel niet waar – 7: Helemaal waar]
10. Ik vond het gemakkelijk om me te concentreren
[1: Geheel niet waar – 7: Helemaal waar]

Tot slot is een aantal vragen uit de studeerbaarheidsvragenlijst (Nadolski & Hummel, 2017) gebruikt om te meten hoe duidelijk de leerlingen de game vinden. Een slechte gebruikservaring verkleint namelijk de kans dat de speler in een *state of flow* komt (Annetta, 2010). Ook een gevoel van controle is hierbij van belang evenals de ervaren tijdsduur.

1. Ik vind dat ik voldoende controle heb binnen de game
[1: Geheel mee oneens – 5: Geheel mee eens]
2. Ik vind de uitleg bij de besturing van de game duidelijk
[1: Geheel mee oneens – 5: Geheel mee eens]
3. Het is voldoende duidelijk wat je in de game moest doen
[1: Geheel mee oneens – 5: Geheel mee eens]
4. Ik vind de uitleg bij de opdrachten in de game duidelijk
[1: Geheel mee oneens – 5: Geheel mee eens]
5. Ik vind dat de game te lang duurt*
[1: Geheel mee oneens – 5: Geheel mee eens]

Interactivity

In gevalideerde vragenlijsten zijn geen vragen opgenomen over de *interactivity*. De onderzoeker heeft onderstaande vragen daarom zelf ontwikkeld.

1. In deze game is voldoende mogelijkheid om vragen te stellen of opmerkingen te maken
[1: Geheel mee oneens – 5: Geheel mee eens]
2. Ik heb het contact dat ik tijdens de game had als prettig ervaren
[1: Geheel mee oneens – 5: Geheel mee eens]
3. Ik zou willen dat ik meer contact kon hebben tijdens het spel*
[1: Geheel mee oneens – 5: Geheel mee eens]

Increasing complexity

De *increasing complexity* werd gemeten aan de hand van de NASA-TLX (Hart & Staveland, 1988) die over de taakbelasting gaat.

1. Hoe mentaal belastend waren de opdrachten in de game?*[
[1: Zeer laag belastend – 7: Zeer hoog belastend]
2. Hoe fysiek belastend waren de opdrachten in de game?*[
[1: Zeer laag belastend – 7: Zeer hoog belastend]
3. Hoe gehaast was het tempo van de opdrachten in de game?*[
[1: Zeer laag belastend – 7: Zeer hoog belastend]
4. Hoe succesvol was je in het doen van de opdrachten in de game?
[1: Zeer laag belastend – 7: Zeer hoog belastend]
5. Hoe hard moest je je inspannen om de opdrachten in de game succesvol te kunnen doen?*[
[1: Zeer laag belastend – 7: Zeer hoog belastend]
6. Hoeveel negatieve gevoelens had je tijdens de opdrachten in de game?*[
[1: Zeer laag belastend – 7: Zeer hoog belastend]

Verder is een tweetal vragen opgenomen uit de studeerbaarheidsvragenlijst (Nadolski & Hummel, 2017) die meten hoe gemakkelijk de game bevonden wordt en of de opdrachten in de goede volgorde staan.

1. Ik vind de game gemakkelijk
[1: Geheel mee oneens – 5: Geheel mee eens]
2. Ik vind dat de opdrachten in de game in een goede volgorde staan
[1: Geheel mee oneens – 5: Geheel mee eens]

Informed teaching

Informed teaching werd gemeten aan de hand van vier vragen uit de studeerbaarheidsvragenlijst (Nadolski & Hummel, 2017). Deze vragen meten de terugkoppeling, aanwijzingen en beoordeling.

1. In de game zouden meer terugkoppeling en aanwijzingen gegeven moeten worden*[
[1: Geheel mee oneens – 5: Geheel mee eens]
2. Ik vind de terugkoppelingen en aanwijzingen in de game goed bruikbaar
[1: Geheel mee oneens – 5: Geheel mee eens]

3. Ik vind dat de ingebouwde begeleiding bij de opdrachten in de game goed is
[1: Geheel mee oneens – 5: Geheel mee eens]
4. Ik wist waarop ik beoordeeld zou worden bij de game
[1: Geheel mee oneens – 5: Geheel mee eens]

Instructional

Tot slot werd *instructional* gemeten met behulp van de studeerbaarheidsvragenlijst (Nadolski & Hummel, 2017). Deze vragen gaan in op de kennis over het vak natuuronderwijs.

1. In de game zou meer kennis over natuuronderwijs verwerkt moeten zijn*
[1: Geheel mee oneens – 5: Geheel mee eens]
2. Voor het uitvoeren van de opdrachten kreeg ik voldoende informatie
[1: Geheel mee oneens – 5: Geheel mee eens]
3. In deze game wordt de kennis over natuuronderwijs voldoende getoetst
[1: Geheel mee oneens – 5: Geheel mee eens]

*Voor deze vragen is de score omgekeerd

Bijlage 4. Kennistesten

Kennistest 1

Datum:

Leerlingnummer:

Vraag 1:

Waar lette de plantendetective Linnaeus op bij het verdelen van planten in groepen?

- a. Hoe planten zich voortplanten
- b. In welke landen planten groeien
- c. Hoe lang een plant in leven blijft

Vraag 2:

Wat hoort niet bij een sporenplant?

- a. Bloemen
- b. Wortels
- c. Bladeren

Vraag 3:

Kunnen sporenplanten zich ook voortplanten?

- a. Ja, doordat ze een stukje van zichzelf achterlaten
- b. Ja, doordat er een bloem uit de plant groeit
- c. Nee

Vraag 4:

Waaruit bestaat het plantenrijk?

- a. Sporenplanten en schimmels
- b. Schimmels en zaadplanten
- c. Sporenplanten en zaadplanten

Vraag 5:

Waar zitten de zaden van een zaadplant verstopt?

- a. In de vruchten
- b. In de wortels
- c. In de bladeren

Vraag 6:

Waar hoort een varen bij?

- a. Zaadplanten
- b. Sporenplanten
- c. Schimmels

Vraag 7:

Waar kunnen schimmels van leven?

- a. Water
- b. Zonlicht
- c. Natuurlijk afval

Vraag 8:

Een perenboom is

- a. Bedektzadig
- b. Naaktzadig

Vraag 9:

In Nederland zijn meer naaktzadige planten dan bedektzadige planten

- a. Waar
- b. Niet waar, het zijn er minder
- c. Niet waar, het zijn er evenveel

Vraag 10:

Wat betekent determineren?

- a. Een plant uit elkaar halen
- b. Het leefgebied van een plant onderzoeken
- c. De onderdelen van een plant bekijken om te weten welke soort het is

Kennistest 2

Datum:

Leerlingnummer:

Vraag 1:

Waar lette de plantendetective Linnaeus niet op bij het verdelen van planten in groepen?

- a. Hoe planten er uit zien
- b. Hoe planten zich voortplanten
- c. Hoeveel bladeren een plant heeft

Vraag 2:

Wat hoort niet bij een sporenplant?

- a. Wortels
- b. Vruchten
- c. Stengels

Vraag 3:

Kunnen sporenplanten zich ook voortplanten?

- a. Nee
- b. Ja, dit doen ze met een spore
- c. Ja, dit doen ze met een zaadje

Vraag 4:

Waaruit bestaat het plantenrijk?

- a. Schimmels en sporenplanten
- b. Zaadplanten en sporenplanten
- c. Zaadplanten en schimmels

Vraag 5:

Waar zitten de zaden van een zaadplant verstopt?

- a. In de stengel
- b. In de bladeren
- c. In de vruchten

Vraag 6:

Waar hoort een appelboom bij?

- a. Zaadplanten
- b. Sporenplanten
- c. Schimmels

Vraag 7:

Waar kunnen schimmels van leven?

- a. Stoffen in de lucht

- b. Dode bladeren
- c. Water

Vraag 8:

Een naaldboom is

- a. Bedektzadig
- b. Naaktzadig

Vraag 9:

In Nederland zijn meer bedektzadige planten dan naaktzadige planten

- a. Waar
- b. Niet waar, het zijn er minder
- c. Niet waar, het zijn er evenveel

Vraag 10:

Wat betekent determineren?

- a. Vragen beantwoorden om zo te ontdekken wat voor plant het is
- b. Zaden van planten verzamelen
- c. Het aantal planten in de omgeving tellen

Kennistest 3

Datum:

Leerlingnummer:

Vraag 1:

Waar lette de plantendetective Linnaeus op bij het verdelen van planten in groepen?

- a. Of een plant eetbaar is
- b. Hoe planten er uit zien
- c. Hoe de wortels van een plant groeien

Vraag 2:

Wat hoort niet bij een sporenplant?

- a. Stengels
- b. Zaden
- c. Bladeren

Vraag 3:

Kunnen sporenplanten zich ook voortplanten?

- a. Ja, hier gebruiken ze hun sporendoosjes voor
- b. Ja, via de vruchten van een sporenplant
- c. Nee

Vraag 4:

Waaruit bestaat het plantenrijk?

- a. Sporenplanten en zaadplanten
- b. Zaadplanten en schimmels
- c. Schimmels en sporenplanten

Vraag 5:

Waar zitten de zaden van een zaadplant verstopt?

- a. In de wortels
- b. In de vruchten
- c. In de stengel

Vraag 6:

Waar hoort een paddenstoel bij?

- a. Zaadplanten
- b. Sporenplanten
- c. Schimmels

Vraag 7:

Waar kunnen schimmels van leven?

- a. Voedingsstoffen uit dieren en planten

- b. Zonlicht
- c. Stoffen in de lucht

Vraag 8:

Een tomatenplant is

- a. Bedektzadig
- b. Naaktzadig

Vraag 9:

In Nederland zijn minder bedektzadige planten dan naaktzadige planten

- a. Waar
- b. Niet waar, het zijn er meer
- c. Niet waar, het zijn er evenveel

Vraag 10:

Wat betekent determineren?

- a. Een plant opmeten
- b. Aanwijzingen zoeken om te achterhalen om welke plantensoort het gaat
- c. Onderzoeken op welke manier een plant zich voortplant